MONOTERPENE SYNTHASES FROM GRAND FIR (*Abies grandis*)

Related Applications

The present application is a continuation-in-part of international application serial number US98/14528, filed on July 10, 1998, which claims benefit of priority from United States provisional application serial number 60/052,249 filed on July 11, 1997. This invention was funded in part by grant GM-3135A from the National Institutes of Health and by grant 97-35302-4432 from the United States Department of Agriculture. The government has certain rights in this invention.

Field of the Invention

The present invention relates to nucleic acid sequences which code for monoterpene synthases from gymnosperm plant species, in particular from Grand fir (*Abies grandis*), including (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, limonene synthase, myrcene synthase, and pinene synthase, to vectors containing the sequences, to host cells containing the sequences, to plant seeds expressing the sequences and to methods of producing recombinant monoterpene synthases and their mutants.

Background of the Invention

Chemical defense of conifer trees against bark beetles and their associated fungal pathogens relies primarily upon constitutive and inducible oleoresin biosynthesis (Johnson, M.A., and Croteau, R. (1987) in *Ecology and Metabolism of Plant Lipids* (Fuller, G., and Nes, W.D., eds.) pp. 76-91, American Chemical Society Symposium Series 325, Washington, DC; Gijzen, M., Lewinsohn, E., Savage, T.J., and Croteau, R.B. (1993) in *Bioactive Volatile Compounds from Plants*

(Teranishi, R., Buttery, R.G., and Sugisawa, H., eds.) pp. 8-22, American Chemical Society Symposium Series 525, Washington, DC). This defensive secretion is a complex mixture of monoterpene and sesquiterpene olefins (turpentine) and diterpene resin acids (rosin) that is synthesized constitutively in the epithelial cells of specialized structures, such as resin ducts and blisters or, in the case of induced oleoresin formation, in undifferentiated cells surrounding wound sites (Lewinsohn, E., Gijzen, M., Savage, T.J., and Croteau, R. (1991) *Plant Physiol.* **96**:38-43). The volatile fraction of conifer oleoresin, which is toxic to both bark beetles and their fungal associates (Raffa, K.F., Berryman, A.A., Simasko, J., Teal, W., and Wong, B.L. (1985) *Environ. Entomol.* **14**:552-556), may consist of up to 30 different monoterpenes (Lewinsohn, E., Savage, T.J., Gijzen, M., and Croteau, R. (1993) *Phytochem. Anal.* **4**:220-225), including acyclic types (e.g., myrcene), monocyclic types (e.g., limonene) and bicyclic types (e.g., pinenes) (FIGURE 1). Although the oleoresin is toxic, many bark beetle species nevertheless employ turpentine volatiles in host selection and can convert various monoterpene components into aggregation or sex pheromones to promote coordinated mass attack of the host (Gijzen, M., Lewinsohn, E., Savage, T.J., and Croteau, R.B. (1993) in *Bioactive Volatile Compounds from Plants* (Teranishi, R., Buttery, R.G., and Sugisawa, H., eds.) pp. 8-22, American Chemical Society Symposium Series 525, Washington, DC; Byers, J.A. (1995) in *Chemical Ecology of Insects 2* (Cardé, R.T., and Bell, W.J., eds.) pp. 154-213, Chapman and Hall, New York). In Grand fir (*Abies grandis*), increased formation of oleoresin monoterpenes, sesquiterpenes and diterpenes is induced by bark beetle attack (Lewinsohn, E., Gijzen, M., Savage, T.J., and Croteau,†R. (1991) *Plant Physiol.* **96**:38-43; Raffa, K.F., and Berryman, A.A. (1982) *Can. Entomol.* **114**:797-810; Lewinsohn, E., Gijzen, M., and Croteau, R. (1991) *Plant Physiol.* **96**:44-49), and this inducible defense response is mimicked by mechanically wounding sapling stems (Lewinsohn, E., Gijzen, M., Savage, T.J., and Croteau, R. (1991) *Plant Physiol.* **96**:38-43; Lewinsohn, E., Gijzen, M., and Croteau, R. (1991) *Plant Physiol.* **96**:44-49; Funk, C., Lewinsohn, E., Stofer Vogel, B., Steele C., and Croteau, R. (1994) *Plant Physiol.* **106**:999-1005). Therefore, Grand fir has been developed as a model system to study the biochemical and molecular genetic regulation of constitutive and inducible terpene biosynthesis in conifers (Steele, C., Lewinsohn, E., and Croteau, R. (1995) *Proc. Natl. Acad. Sci. USA* **92**:4164-4168).

Most monoterpenes are derived from geranyl diphosphate, the ubiquitous C_{10} intermediate of the isoprenoid pathway, by synthases which catalyze the divalent metal ion-dependent ionization (to 1, FIGURE 1) and isomerization of this substrate to enzyme-bound linalyl diphosphate which, following rotation about C2-C3, undergoes a second ionization (to 2, FIGURE 1) followed by cyclization to the α-terpinyl cation, the first cyclic intermediate en route to both monocyclic and bicyclic products (Croteau, R., and Cane, D.E. (1985) *Methods Enzymol.* **110**:383-405; Croteau, R. (1987) *Chem. Rev.* **87**:929-954) (FIGURE 1). Acyclic monoterpenes, such as myrcene, may arise by deprotonation of carbocations 1 or 2, whereas the isomerization step to linalyl diphosphate is required in the case of cyclic types, such as limonene and pinenes, which cannot be derived from geranyl diphosphate directly because of the geometric impediment of the *trans*-double bond at C2-C3 (Croteau, R., and Cane, D.E. (1985) *Methods Enzymol.* **110**:383-405; Croteau, R. (1987) *Chem. Rev.* **87**:929-954). Many monoterpene synthases catalyze the formation of multiple products, including acyclic, monocyclic and bicyclic types, by variations on this basic mechanism (Gambliel, H., and Croteau, R. (1984) *J. Biol. Chem.* **259**:740-748; Croteau, R., Satterwhite, D.M., Cane, D.E., and Chang, C.C. (1988) *J. Biol. Chem.* **263**:10063-10071; Croteau, R., and Satterwhite, D.M. (1989) *J. Biol. Chem.* **264**:15309-15315). For example, (-)-limonene synthase, the principal monoterpene synthase of spearmint (*Mentha spicata*) and peppermint (*M.* x *piperita*), produces small amounts of myrcene, (-)-α-pinene and (-)-β-pinene in addition to the monocyclic product (Rajaonarivony, J.I.M., Gershenzon, J., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **296**:49-57; Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024. Conversely, six different inducible monoterpene synthase activities have been demonstrated in extracts of wounded Grand fir stem (Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**:267-273) indicating that formation of acyclic, monocyclic and bicyclic monoterpenes in this species involves several genes encoding distinct catalysts. The inducible (-)-pinene synthase has been purified (Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173), and isotopically sensitive branching experiments employed to demonstrate that this enzyme synthesizes both (-)-α- and (-)-β-pinene (Wagschal, K., Savage, T.J., and Croteau, R. (1991) *Tetrahedron* **47**:5933-5944).

Deciphering the molecular genetic control of oleoresinosis and examining structure-function relationships among the monoterpene synthases of Grand fir

requires isolation of the cDNA species encoding these key enzymes. Although a protein-based cloning strategy was recently employed to acquire a cDNA for the major wound-inducible diterpene synthase from Grand fir, abietadiene synthase (Funk, C., Lewinsohn, E., Stofer Vogel, B., Steele C., and Croteau, R. (1994) *Plant Physiol.* **106**:999-1005; LaFever, R.E., Stofer Vogel, B., and Croteau, R. (1994) *Arch. Biochem. Biophys.* **313**:139-149; Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268), all attempts at the reverse genetic approach to cloning of Grand fir monoterpene synthases have failed (Steele, C., Lewinsohn, E., and Croteau, R. (1995) *Proc. Natl. Acad. Sci. USA* **92**:4164-4168). As an alternative, a similarity-based PCR strategy was developed (Steele, C., Lewinsohn, E., and Croteau, R. (1995) *Proc. Natl. Acad. Sci. USA* **92**:4164-4168) that employed sequence information from terpene synthases of angiosperm origin, namely a monoterpene synthase, (-)-4*S*-limonene synthase, from spearmint (*Mentha spicata*, Lamiaceae) (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024), a sesquiterpene synthase, 5-*epi*-aristolochene synthase, from tobacco (*Nicotiana tabacum*, Solanaceae) (Facchini, P.J., and Chappell, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**:11088-11092), and a diterpene synthase, casbene synthase, from castor bean (*Ricinus communis*, Euphorbiaceae) (Mau, C.J.D., and West, C.A. (1994) *Proc. Natl. Acad. Sci. USA* **91**:8497-8501).

Monoterpenes have significant potential for cancer prevention and treatment. Monoterpenes such as limonene, perillyl alcohol, carvone, geraniol and farnesol not only reduce tumor incidence and slow tumor proliferation, but have also been reported to cause regression of established solid tumors by initiating apoptosis (Mills J.J., Chari R.S., Boyer I.J., Gould M.N., Jirtle R.L., *Cancer Res.*, **55**:979-983, 1995). Terpenes have activity against cancers such as mammary, colon, and prostate. Clinical trials are being pursued (Seachrist L, *J. NIH Res.* **8**:43) in patients with various types of advanced cancers to validate the health benefits of dietary terpenes for humans. However, terpenes are present in Western diets at levels that are probably inadequate for any significant preventive health benefits. Daily supplementation of the diet with a terpene concentrate (10-20 g/day) would appear to be the most rational strategy for dietary therapy of diagnosed cases of cancer. This invention envisages the production of such nutritionally beneficial terpenes in, for example, vegetable oils consumed daily via the engineering of relevant genes from

Grand fir into oil seed crop plants such as oil seed brassica (canola), soybean and corn.

Brief Description of the Drawings

The foregoing aspects and many of the attendant advantages of this invention will become more readily appreciated as the same becomes better understood by reference to the following detailed description, when taken in conjunction with the accompanying drawings, wherein:

FIGURE 1 is a schematic representation depicting the mechanism for the conversion of geranyl diphosphate to myrcene, (-)-limonene, β-phellandrene, (-)-α-pinene and (-)-β-pinene by monoterpene synthases from Grand fir. Formation of the monocyclic and bicyclic products requires preliminary isomerization of geranyl diphosphate to linalyl diphosphate. The acyclic product could be formed from either geranyl diphosphate or linalyl diphosphate via carbocations 1 or 2. *OPP* denotes the diphosphate moiety.

FIGURE 2 is a sequence comparison of plant terpene synthases. A three-letter designation (*Tps*) for the gene family is proposed with sub-groups (*Tpsa* through *Tpsf*) defined by a minimum of 40 % amino acid identity between members.

FIGURE 3 depicts a GLC-MS analysis of the products of the recombinant protein encoded by *AG2.2* (SEQ ID NO:1), the sequence of the protein encoded by clone *AG2.2* being set forth in SEQ ID NO:2. The GLC profile of the total pentane-soluble products generated from geranyl diphosphate when incubated with a cell-free extract of *E. coli* XL1-Blue/pGAG2.2 is illustrated (FIGURE 3A), as are the mass fragmentation patterns for the monoterpene product with R_t = 12.22 min (FIGURE 3B) and for authentic myrcene (FIGURE 3C).

FIGURE 4 depicts a GLC-MS analysis of the products of the recombinant protein encoded by *AG3.18* (SEQ ID NO:3), the sequence of the protein encoded by clone *AG3.18* (SEQ ID NO:3) being set forth in SEQ ID NO:4. The GLC profile of the total pentane-soluble products generated from geranyl diphosphate when incubated with a cell-free extract of *E. coli* XL1-Blue/pGAG3.18 is illustrated (FIGURE 4A), as are the mass fragmentation patterns (selected ion mode) for the monoterpene products with R_t = 11.34 min (FIGURE 4B), and R_t = 13.37 min (FIGURE 4D), and for authentic α-pinene (FIGURE 4C) and authentic β-pinene (FIGURE 4E).

FIGURE 5 depicts a GLC-MS analysis of the products of the recombinant protein encoded by *AG10* (SEQ ID NO:5), the sequence of the protein encoded by

clone *AG10* (SEQ ID NO:5) being set forth in SEQ ID NO:6. The GLC profile of the total pentane-soluble products generated from geranyl diphosphate when incubated with a cell-free extract of *E. coli* BL21(DE3)/pSBAG10 is illustrated (FIGURE 5A), as are the mass fragmentation patterns for the principal monoterpene product with R_t = 13.93 min (FIGURE 5B) and for authentic limonene (FIGURE 5C).

FIGURE 6A depicts a total ion chromatogram of monoterpene products derived from geranyl diphosphate by a (-)-camphene synthase of the invention.

FIGURE 6B depicts the mass spectrum and retention time for the principal enzyme product shown in FIGURE 6A.

FIGURE 6C depicts the mass spectrum and retention time for the authentic camphene standard.

FIGURE 7A depicts a total ion chromatogram of monoterpene products derived from geranyl diphosphate by a (-)-β-phellandrene synthase of the invention.

FIGURE 7B depicts the mass spectrum and retention time for the principal enzyme product shown in FIGURE 7A.

FIGURE 7C depicts the mass spectrum and retention time for the authentic β-phellandrene standard.

FIGURE 8A depicts a total ion chromatogram of monoterpene products derived from geranyl diphosphate by a terpinolene synthase of the invention.

FIGURE 8B depicts the mass spectrum and retention time for the principal enzyme product shown in FIGURE 8A.

FIGURE 8C depicts the mass spectrum and retention time for the authentic terpinolene standard.

FIGURE 9A depicts a total ion chromatogram of monoterpene products derived from geranyl diphosphate by a (-)-limonene/(-)-α-pinene synthase of the invention.

FIGURE 9B depicts the mass spectrum and retention time for the principal enzyme product shown in FIGURE 9A.

FIGURE 9C depicts the mass spectrum and retention time for the authentic α-pinene standard.

Summary of the Invention

In accordance with the foregoing, cDNAs encoding (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, myrcene synthase, (-)-limonene synthase and (-)-pinene synthase from Grand fir (*Abies grandis*) have been isolated and sequenced, and the corresponding amino acid

sequences have been deduced. Accordingly, the present invention relates to isolated DNA sequences which code for the expression of monoterpene synthases, including (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, myrcene synthase, (-)-limonene synthase and (-)-pinene synthase, and to isolated nucleic acid molecules that hybridize to portions of Grand fir monoterpene synthase cDNAs, as described more fully herein. In another aspect, the present invention relates to isolated monoterpene synthases, including isolated (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase and (-)-limonene/(-)-α-pinene synthase. In other aspects, the present invention is directed to replicable recombinant cloning vehicles comprising a nucleic acid sequence, e.g., a DNA sequence which codes for a monoterpene synthase such as (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, myrcene synthase, (-)-limonene synthase or (-)-pinene synthase, or for a base sequence sufficiently complementary to at least a portion of DNA or RNA encoding a monoterpene synthase such as (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, myrcene synthase, (-)-limonene synthase or (-)-pinene synthase to enable hybridization therewith (e.g., antisense RNA or fragments of DNA complementary to a portion of DNA or RNA molecules encoding (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, myrcene synthase, (-)-limonene synthase or (-)-pinene synthase which are useful as polymerase chain reaction primers or as probes for any of the foregoing synthases or related genes). In yet other aspects of the invention, modified host cells are provided that have been transformed, transfected, infected and/or injected with a recombinant cloning vehicle and/or DNA sequence of the invention. Thus, the present invention provides for the recombinant expression of (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, myrcene synthase, (-)-limonene synthase and (-)-pinene synthase, and the inventive concepts may be used to facilitate the production, isolation and purification of significant quantities of recombinant (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, myrcene synthase, (-)-limonene synthase and (-)-pinene synthase (or of their primary enzyme products) for subsequent use, to obtain expression or enhanced expression of (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, myrcene synthase, (-)-limonene synthase and (-)-pinene synthase in microorganisms,

animals or plants (including, but not limited to, Brassica, cotton, soybean, safflower, sunflower, coconut, palm, wheat, barley, rice, corn, oats, amaranth, pumpkin, squash, sesame, poppy, grape, mung beans, peanut, peas, beans, broad beans, chick peas, lentils, radish, alfalfa, cocoa, coffee, tree nuts, spinach, culinary herbs, berries, stone fruit and citrus), or may be otherwise employed in an environment where the regulation or expression of the foregoing monoterpene synthases is desired for the production of these synthases, or their enzyme products, or derivatives thereof. In another aspect, the present invention relates to manipulation of monoterpene production to enhance resistance to insects and/or accumulate nutritionally beneficial monoterpenes in oil seeds (such as seeds from Brassica, cotton, soybean, safflower, sunflower, coconut, palm, wheat, barley, rice, corn, oats, amaranth, pumpkin, squash, sesame, poppy, grape, mung beans, peanut, peas, beans, broad beans, chick peas, lentils, radish, alfalfa, cocoa, coffee and tree nuts) and other food stuffs.

Detailed Description of the Preferred Embodiment

As used herein, the terms "amino acid" and "amino acids" refer to all naturally occurring L-α-amino acids or their residues. The amino acids are identified by either the single-letter or three-letter designations:

Asp	D	aspartic acid	Ile	I	isoleucine
Thr	T	threonine	Leu	L	leucine
Ser	S	serine	Tyr	Y	tyrosine
Glu	E	glutamic acid	Phe	F	phenylalanine
Pro	P	proline	His	H	histidine
Gly	G	glycine	Lys	K	lysine
Ala	A	alanine	Arg	R	arginine
Cys	C	cysteine	Trp	W	tryptophan
Val	V	valine	Gln	Q	glutamine
Met	M	methionine	Asn	N	asparagine

As used herein, the term "nucleotide" means a monomeric unit of DNA or RNA containing a sugar moiety (pentose), a phosphate and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of pentose) and that combination of base and sugar is called a nucleoside. The base characterizes the nucleotide with the four bases of DNA being adenine ("A"), guanine ("G"), cytosine ("C") and thymine ("T"). Inosine ("I") is a synthetic base that can be used to substitute for any of the four, naturally-occurring bases (A, C, G or T). The four RNA bases are A,G,C and uracil ("U"). The nucleotide sequences described

herein comprise a linear array of nucleotides connected by phosphodiester bonds between the 3' and 5' carbons of adjacent pentoses.

The term "percent identity" means the percentage of amino acids or nucleotides that occupy the same relative position when two amino acid sequences, or two nucleic acid sequences are aligned side by side.

The term "percent similarity" is a statistical measure of the degree of relatedness of two compared protein sequences. The percent similarity is calculated by a computer program that assigns a numerical value to each compared pair of amino acids based on chemical similarity (e.g., whether the compared amino acids are acidic, basic, hydrophobic, aromatic, etc.) and/or evolutionary distance as measured by the minimum number of base pair changes that would be required to convert a codon encoding one member of a pair of compared amino acids to a codon encoding the other member of the pair. Calculations are made after a best fit alignment of the two sequences have been made empirically by iterative comparison of all possible alignments. (Henikoff, S. and Henikoff, J.G., *Proc. Nat'l. Acad. Sci. USA* **89**:10915-10919, 1992).

"Oligonucleotide" refers to short length single or double stranded sequences of deoxyribonucleotides linked via phosphodiester bonds. The oligonucleotides are chemically synthesized by known methods and purified, for example, on polyacrylamide gels.

The term "myrcene synthase" is used herein to mean an enzyme capable of generating multiple monoterpenes from geranyl diphosphate. The principal and characteristic monoterpene synthesized by myrcene synthase is myrcene, which constitutes at least about 50% of the monoterpene mixture synthesized by myrcene synthase from geranyl diphosphate.

The term "(-)-limonene synthase" is used herein to mean an enzyme capable of generating multiple monoterpenes from geranyl diphosphate. The principal and characteristic monoterpene synthesized by (-)-limonene synthase is (-)-limonene, which constitutes at least about 60% of the monoterpene mixture synthesized by (-)-limonene synthase from geranyl diphosphate.

The term "(-)-pinene synthase" is used herein to mean an enzyme capable of generating multiple monoterpenes from geranyl diphosphate. The principal and characteristic monoterpene synthesized by (-)-pinene synthase is (-)-pinene, which comprises at least about 50% of the monoterpene mixture synthesized by (-)-pinene synthase from geranyl diphosphate.

The term "(-)-camphene synthase" is used herein to mean an enzyme capable of generating multiple monoterpenes from geranyl diphosphate. The principal and characteristic monoterpene synthesized by (-)-camphene synthase is (-)-camphene, which comprises at least about 50% of the monoterpene mixture synthesized by (-)-camphene synthase from geranyl diphosphate.

The term "(-)-β-phellandrene synthase" is used herein to mean an enzyme capable of generating multiple monoterpenes from geranyl diphosphate. The principal and characteristic monoterpene synthesized by (-)-β-phellandrene synthase is (-)-β-phellandrene, which comprises at least about 50% of the monoterpene mixture synthesized by (-)-β-phellandrene synthase from geranyl diphosphate.

The term "terpinolene synthase" is used herein to mean an enzyme capable of generating multiple monoterpenes from geranyl diphosphate. The principal and characteristic monoterpene synthesized by terpinolene synthase is terpinolene which comprises at least about 40% of the monoterpene mixture synthesized by terpinolene synthase from geranyl diphosphate.

The term (-)-limonene/(-)-α-pinene synthase is used herein to mean an enzyme capable of generating multiple monoterpenes from geranyl diphosphate. The principal and characteristic monoterpenes synthesized by (-)-limonene/(-)-α-pinene synthase are (-)-limonene and (-)-α-pinene which comprise at least about 35% and 25%, respectively, of the monoterpene mixture synthesized by (-)-limonene/(-)-α-pinene synthase from geranyl diphosphate.

The abbreviation "SSPE" refers to a buffer used in nucleic acid hybridization solutions. The 20X (twenty times concentrate) stock SSPE buffer solution is prepared as follows: dissolve 175.3 grams of NaCl, 27.6 grams of $NaH_2PO_4H_2O$ and 7.4 grams of EDTA in 800 millilitres of H_2O. Adjust the pH to pH 7.4 with NaOH. Adjust the volume to one liter with H_2O.

The abbreviation "SSC" refers to a buffer used in nucleic acid hybridization solutions. One liter of the 20X (twenty times concentrate) stock SSC buffer solution (pH 7.0) contains 175.3 g sodium chloride and 88.2 g sodium citrate.

The terms "alteration", "amino acid sequence alteration", "variant" and "amino acid sequence variant" refer to monoterpene synthase molecules with some differences in their amino acid sequences as compared to the corresponding, native, i.e., naturally-occurring, monoterpene synthases. Ordinarily, the variants will possess at least about 70% homology with the corresponding native monoterpene synthases, and preferably, they will be at least about 80% homologous with the corresponding, native

monoterpene synthases. The amino acid sequence variants of the monoterpene synthases falling within this invention possess substitutions, deletions, and/or insertions at certain positions. Sequence variants of monoterpene synthases may be used to attain desired enhanced or reduced enzymatic activity, modified regiochemistry or stereochemistry, or altered substrate utilization or product distribution.

Substitutional monoterpene synthase variants are those that have at least one amino acid residue in the native monoterpene synthase sequence removed and a different amino acid inserted in its place at the same position. The substitutions may be single, where only one amino acid in the molecule has been substituted, or they may be multiple, where two or more amino acids have been substituted in the same molecule. Substantial changes in the activity of the monoterpene synthase molecules of the present invention may be obtained by substituting an amino acid with a side chain that is significantly different in charge and/or structure from that of the native amino acid. This type of substitution would be expected to affect the structure of the polypeptide backbone and/or the charge or hydrophobicity of the molecule in the area of the substitution.

Moderate changes in the activity of the monoterpene synthase molecules of the present invention would be expected by substituting an amino acid with a side chain that is similar in charge and/or structure to that of the native molecule. This type of substitution, referred to as a conservative substitution, would not be expected to substantially alter either the structure of the polypeptide backbone or the charge or hydrophobicity of the molecule in the area of the substitution.

Insertional monoterpene synthase variants are those with one or more amino acids inserted immediately adjacent to an amino acid at a particular position in the native monoterpene synthase molecule. Immediately adjacent to an amino acid means connected to either the α-carboxy or α-amino functional group of the amino acid. The insertion may be one or more amino acids. Ordinarily, the insertion will consist of one or two conservative amino acids. Amino acids similar in charge and/or structure to the amino acids adjacent to the site of insertion are defined as conservative. Alternatively, this invention includes insertion of an amino acid with a charge and/or structure that is substantially different from the amino acids adjacent to the site of insertion.

Deletional variants are those where one or more amino acids in the native monoterpene synthase molecules have been removed. Ordinarily, deletional variants

will have one or two amino acids deleted in a particular region of the monoterpene synthase molecule.

The terms "biological activity", "biologically active", "activity" and "active" refer to the ability of the monoterpene synthases of the present invention to convert geranyl diphosphate to a group of monoterpenes, of which myrcene is the principal and characteristic monoterpene synthesized by myrcene synthase, (-)-limonene is the principal and characteristic monoterpene synthesized by (-)-limonene synthase, (-)-pinene is the principal and characteristic monoterpene synthesized by (-)-pinene synthase, (-)-camphene is the principal and characteristic monoterpene synthesized by (-)-camphene synthase, (-)-β-phellandrene is the principal and characteristic monoterpene synthesized by (-)-β-phellandrene synthase, terpinolene is the principal and characteristic monoterpene synthesized by terpinolene synthase, and (-)-limonene and (-)-α-pinene are the principal and characteristic monoterpenes synthesized by (-)-limonene/(-)-α-pinene synthase. The monoterpenes produced by the monoterpene synthases of the present invention are as measured in an enzyme activity assay, such as the assay described in Example 3. Amino acid sequence variants of the monoterpene synthases of the present invention may have desirable altered biological activity including, for example, altered reaction kinetics, substrate utilization product distribution or other characteristics such as regiochemistry and stereochemistry.

The terms "DNA sequence encoding", "DNA encoding" and "nucleic acid encoding" refer to the order or sequence of deoxyribonucleotides along a strand of deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along the translated polypeptide chain. The DNA sequence thus codes for the amino acid sequence.

The terms "replicable expression vector" and "expression vector" refer to a piece of DNA, usually double-stranded, which may have inserted into it a piece of foreign DNA. Foreign DNA is defined as heterologous DNA, which is DNA not naturally found in the host. The vector is used to transport the foreign or heterologous DNA into a suitable host cell. Once in the host cell, the vector can replicate independently of or coincidental with the host chromosomal DNA, and several copies of the vector and its inserted (foreign) DNA may be generated. In addition, the vector contains the necessary elements that permit translating the foreign DNA into a polypeptide. Many molecules of the polypeptide encoded by the foreign DNA can thus be rapidly synthesized.

The terms "transformed host cell," "transformed" and "transformation" refer to the introduction of DNA into a cell. The cell is termed a "host cell", and it may be a prokaryotic or a eukaryotic cell. Typical prokaryotic host cells include various strains of *E. coli*. Typical eukaryotic host cells are plant cells, such as maize cells, yeast cells, insect cells or animal cells. The introduced DNA is usually in the form of a vector containing an inserted piece of DNA. The introduced DNA sequence may be from the same species as the host cell or from a different species from the host cell, or it may be a hybrid DNA sequence, containing some foreign DNA and some DNA derived from the host species.

In accordance with the present invention, cDNAs encoding myrcene synthase (SEQ ID NO:1), (-)-pinene synthase (SEQ ID NO:3) and (-)-limonene synthase (SEQ ID NO:5) from Grand fir (*Abies grandis*) were isolated and sequenced in the following manner. Based on comparison of sequences of limonene synthase from spearmint (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024), 5-*epi*-aristolochene synthase from tobacco (Facchini, P.J., and Chappell, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**:11088-11092), and casbene synthase from castor bean (Mau, C.J.D., and West, C.A. (1994) *Proc. Natl. Acad. Sci. USA* **91**:8497-8501), four conserved regions were identified for which a set of consensus, degenerate primers were synthesized: Primer A (SEQ ID NO:7), Primer B (SEQ ID NO:8), Primer C (SEQ ID NO:9) and Primer D (SEQ ID NO:10). Primers A (SEQ ID NO:7), B (SEQ ID NO:8), and D (SEQ ID NO:10) were sense primers, while Primer C (SEQ ID NO:9), was an antisense primer. Each of the sense primers, A (SEQ ID NO:7), B (SEQ ID NO:8) and D (SEQ ID NO:10), was used for PCR in combination with antisense primer C (SEQ ID NO:9) by employing a broad range of amplification conditions. Analysis of the PCR reaction products by agarose gel electrophoresis revealed that only the combination of primers C (SEQ ID NO:9) and D (SEQ ID NO:10) generated a specific PCR product of approximately 110 bps.

The 110 bps PCR product was gel purified, ligated into a plasmid, and transformed into *E. coli* XL1-Blue cells. Plasmid DNA was prepared from 41 individual transformants and the inserts were sequenced. Four different insert sequences were identified, and were designated as probes 1 (SEQ ID NO:11), 2 (SEQ ID NO:12), 4 (SEQ ID NO:13) and 5 (SEQ ID NO:14). Probes 1 (SEQ ID NO:11), 2 (SEQ ID NO:12), 4 (SEQ ID NO:13) and 5 (SEQ ID NO:14) were used to screen a cDNA library made from mRNA extracted from wounded Grand fir stems,

and the longest clone that hybridized to each of these probes was isolated and sequenced. Thus, clone *AG1.28* (SEQ ID NO:15) is the longest cDNA clone that hybridized to probe 1 (SEQ ID NO:11), clone *AG2.2* (SEQ ID NO:1) is the longest cDNA clone that hybridized to probe 2 (SEQ ID NO:12), clone *AG4.30* (SEQ ID NO:17) is the longest cDNA clone that hybridized to probe 4 (SEQ ID NO: 13), and clone *AG5.9* (SEQ ID NO:19) is the longest cDNA clone that hybridized to probe 5 (SEQ ID NO:14).

Truncated clone *AG1.28* (SEQ ID NO:15) resembled most closely in size and sequence (72% similarity, 49% identity) a diterpene cyclase, abietadiene synthase, from Grand fir. Clones *AG4.30* (SEQ ID NO:17) and *AG5.9* (SEQ ID NO:19) encode sesquiterpene synthases. Sequence and functional analysis of clone *AG2.2* (SEQ ID NO:1) revealed that it encoded the monoterpene synthase, myrcene synthase.

Alignment of the four new terpene synthase cDNA sequences *AG1.28* (SEQ ID NO:15), *AG2.2* (SEQ ID NO:1), *AG4.30* (SEQ ID NO:17) and *AG5.9* (SEQ ID NO:19) with that for abietadiene synthase (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* 271:23262-23268) allowed the identification of several, conserved sequence motifs. Two new sense PCR primers, primer E (SEQ ID NO:21) and primer F (SEQ ID NO:22) were designed based on the sequence of the conserved protein sequence motifs. A new antisense PCR primer, primer G (SEQ ID NO:23), was designed based on limited sequence information available from pinene synthase. The combination of primer E (SEQ ID NO:21) and primer G (SEQ ID NO:23) amplified a cDNA product of 1022 bps, which was designated as probe 3 (SEQ ID NO:24).

Probe 3 (SEQ ID NO:24) was used to screen a cDNA library made from mRNA extracted from wounded Grand fir stems. Hybridization of 10^5 Grand fir λZAP II cDNA clones with probe 3 (SEQ ID NO:24) yielded two types of signals comprised of about 400 strongly positive clones and an equal number of weak positives, indicating that the probe recognized more than one type of cDNA. Thirty-four of the former clones and eighteen of the latter were purified, the inserts were selected by size (2.0-2.5 kb), and the *in vivo* excised clones were partially sequenced from both ends. Those clones which afforded weak hybridization signals were shown to contain inserts that were either identical to myrcene synthase clone *AG2.2* (SEQ ID NO:1) or exhibited no significant sequence similarity to terpene synthases.

Clones which gave strong hybridization signals segregated into distinct sequence groups represented by clone *AG3.18* (SEQ ID NO:3) and clone *AG10* (SEQ ID NO:5). Both *AG3.18* (SEQ ID NO:3) and *AG10* (SEQ ID NO:5) were subcloned into plasmid expression vectors and expressed in *E. coli*. When extracts of the induced cells were tested for terpene synthase activity with all of the potential prenyl diphosphate substrates, only geranyl diphosphate was utilized. Extracts from *E. coli* containing the *AG10* (SEQ ID NO:5) expression construct converted geranyl diphosphate to the (-)-4S enantiomer of limonene as the major product, indicating that *AG10* (SEQ ID NO:5) encodes (-)-limonene synthase. Similar analysis of the monoterpene products generated from geranyl diphosphate by cell-free extracts of *E. coli* containing the *AG3.18* (SEQ ID NO:3) insert ligated into an expression vector revealed the presence of a 42:58% mixture of α-pinene and β-pinene, the same product ratio previously described for the purified, native (-)-pinene synthase from Grand fir. These data indicate that *AG3.18* (SEQ ID NO:3) encodes (-)-pinene synthase.

Additionally, cDNA molecules encoding (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase were isolated and characterized as described in Example 11.

The isolation of cDNAs encoding (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase and myrcene synthase permits the development of efficient expression systems for these functional enzymes; provides useful tools for examining the developmental regulation of monoterpene biosynthesis; permits investigation of the reaction mechanism(s) of these unusual, multiproduct enzymes, and permits the isolation of other monoterpene synthases including (-)-camphene synthases, (-)-β-phellandrene synthases, terpinolene synthases, (-)-limonene/(-)-α-pinene synthases, (-)-limonene synthases, (-)-pinene synthases and myrcene synthases. The isolation of the (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase and myrcene synthase cDNAs also permits the transformation of a wide range of organisms in order to introduce monoterpene biosynthesis *de novo*, or to modify endogenous monoterpene biosynthesis.

Substitution of the presumptive targeting sequence of the cloned monoterpene synthases (*e.g.*, SEQ ID NO:2, amino acids 1 to 61; SEQ ID NO:4, amino acids 1 to 61; SEQ ID NO:6, amino acids 1 to 66) with other transport sequences well known in

the art (see, e.g., von Heijne et al., *Eur. J. Biochem.* **180**:535-545, 1989; Stryer, *Biochemistry*, W.H. Freeman and Company, New York, NY, p. 769 [1988]) may be employed to direct the cloned monoterpene synthases of the invention to other cellular or extracellular locations.

In addition to the native monoterpene synthase amino acid sequences, sequence variants produced by deletions, substitutions, mutations and/or insertions are intended to be within the scope of the invention except insofar as limited by the prior art. The monoterpene synthase amino acid sequence variants of this invention may be constructed by mutating the DNA sequences that encode the wild-type synthases, such as by using techniques commonly referred to as site-directed mutagenesis. Nucleic acid molecules encoding the monoterpene synthases of the present invention can be mutated by a variety of PCR techniques well known to one of ordinary skill in the art. *See, e.g.*, "PCR Strategies", M.A. Innis, D.H. Gelfand and J.J. Sninsky, eds., 1995, Academic Press, San Diego, CA (Chapter 14); "PCR Protocols: A Guide to Methods and Applications", M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White, eds., Academic Press, NY (1990).

By way of non-limiting example, the two primer system utilized in the Transformer Site-Directed Mutagenesis kit from Clontech, may be employed for introducing site-directed mutants into the monoterpene synthase genes of the present invention. Following denaturation of the target plasmid in this system, two primers are simultaneously annealed to the plasmid; one of these primers contains the desired site-directed mutation, the other contains a mutation at another point in the plasmid resulting in elimination of a restriction site. Second strand synthesis is then carried out, tightly linking these two mutations, and the resulting plasmids are transformed into a *mutS* strain of *E. coli*. Plasmid DNA is isolated from the transformed bacteria, restricted with the relevant restriction enzyme (thereby linearizing the unmutated plasmids), and then retransformed into *E. coli*. This system allows for generation of mutations directly in an expression plasmid, without the necessity of subcloning or generation of single-stranded phagemids. The tight linkage of the two mutations and the subsequent linearization of unmutated plasmids results in high mutation efficiency and allows minimal screening. Following synthesis of the initial restriction site primer, this method requires the use of only one new primer type per mutation site. Rather than prepare each positional mutant separately, a set of "designed degenerate" oligonucleotide primers can be synthesized in order to introduce all of the desired mutations at a given site simultaneously. Transformants can be screened by

sequencing the plasmid DNA through the mutagenized region to identify and sort mutant clones. Each mutant DNA can then be restricted and analyzed by electrophoresis on Mutation Detection Enhancement gel (J.T. Baker) to confirm that no other alterations in the sequence have occurred (by band shift comparison to the unmutagenized control).

The verified mutant duplexes in the pET (or other) overexpression vector can be employed to transform *E. coli* such as strain *E. coli* BL21(DE3)pLysS, for high level production of the mutant protein, and purification by standard protocols. The method of FAB-MS mapping can be employed to rapidly check the fidelity of mutant expression. This technique provides for sequencing segments throughout the whole protein and provides the necessary confidence in the sequence assignment. In a mapping experiment of this type, protein is digested with a protease (the choice will depend on the specific region to be modified since this segment is of prime interest and the remaining map should be identical to the map of unmutagenized protein). The set of cleavage fragments is fractionated by microbore HPLC (reversed phase or ion exchange, again depending on the specific region to be modified) to provide several peptides in each fraction, and the molecular weights of the peptides are determined by FAB-MS. The masses are then compared to the molecular weights of peptides expected from the digestion of the predicted sequence, and the correctness of the sequence quickly ascertained. Since this mutagenesis approach to protein modification is directed, sequencing of the altered peptide should not be necessary if the MS agrees with prediction. If necessary to verify a changed residue, CAD-tandem MS/MS can be employed to sequence the peptides of the mixture in question, or the target peptide purified for subtractive Edman degradation or carboxypeptidase Y digestion depending on the location of the modification.

In the design of a particular site directed mutagenesis, it is generally desirable to first make a non-conservative substitution (e.g., Ala for Cys, His or Glu) and determine if activity is greatly impaired as a consequence. The properties of the mutagenized protein are then examined with particular attention to the kinetic parameters of K_m and k_{cat} as sensitive indicators of altered function, from which changes in binding and/or catalysis *per se* may be deduced by comparison to the native enzyme. If the residue is by this means demonstrated to be important by activity impairment, or knockout, then conservative substitutions can be made, such as Asp for Glu to alter side chain length, Ser for Cys, or Arg for His. For hydrophobic segments, it is largely size that is usefully altered, although aromatics can

also be substituted for alkyl side chains. Changes in the normal product distribution can indicate which step(s) of the reaction sequence have been altered by the mutation. Modification of the hydrophobic pocket can be employed to change binding conformations for substrates and result in altered regiochemistry and/or stereochemistry.

Other site directed mutagenesis techniques may also be employed with the nucleotide sequences of the invention. For example, restriction endonuclease digestion of DNA followed by ligation may be used to generate deletion variants of (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase and myrcene synthase, as described in section 15.3 of Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, New York, NY [1989]). A similar strategy may be used to construct insertion variants, as described in section 15.3 of Sambrook et al., *supra*.

Oligonucleotide-directed mutagenesis may also be employed for preparing substitution variants of this invention. It may also be used to conveniently prepare the deletion and insertion variants of this invention. This technique is well known in the art as described by Adelman et al. (*DNA* 2:183 [1983]); Sambrook et al., *supra*; "Current Protocols in Molecular Biology", 1991, Wiley (NY), F.T. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.D. Seidman, J.A. Smith and K. Struhl, eds.

Generally, oligonucleotides of at least 25 nucleotides in length are used to insert, delete or substitute two or more nucleotides in the nucleic acid molecules encoding monoterpene synthases of the invention. An optimal oligonucleotide will have 12 to 15 perfectly matched nucleotides on either side of the nucleotides coding for the mutation. To mutagenize nucleic acids encoding wild-type monoterpene synthases of the invention, the oligonucleotide is annealed to the single-stranded DNA template molecule under suitable hybridization conditions. A DNA polymerizing enzyme, usually the Klenow fragment of *E. coli* DNA polymerase I, is then added. This enzyme uses the oligonucleotide as a primer to complete the synthesis of the mutation-bearing strand of DNA. Thus, a heteroduplex molecule is formed such that one strand of DNA encodes the wild-type synthase inserted in the vector, and the second strand of DNA encodes the mutated form of the synthase inserted into the same vector. This heteroduplex molecule is then transformed into a suitable host cell.

Mutants with more than one amino acid substituted may be generated in one of several ways. If the amino acids are located close together in the polypeptide

chain, they may be mutated simultaneously using one oligonucleotide that codes for all of the desired amino acid substitutions. If however, the amino acids are located some distance from each other (separated by more than ten amino acids, for example) it is more difficult to generate a single oligonucleotide that encodes all of the desired changes. Instead, one of two alternative methods may be employed. In the first method, a separate oligonucleotide is generated for each amino acid to be substituted. The oligonucleotides are then annealed to the single-stranded template DNA simultaneously, and the second strand of DNA that is synthesized from the template will encode all of the desired amino acid substitutions. An alternative method involves two or more rounds of mutagenesis to produce the desired mutant. The first round is as described for the single mutants: wild-type monoterpene synthase DNA is used for the template, an oligonucleotide encoding the first desired amino acid substitution(s) is annealed to this template, and the heteroduplex DNA molecule is then generated. The second round of mutagenesis utilizes the mutated DNA produced in the first round of mutagenesis as the template. Thus, this template already contains one or more mutations. The oligonucleotide encoding the additional desired amino acid substitution(s) is then annealed to this template, and the resulting strand of DNA now encodes mutations from both the first and second rounds of mutagenesis. This resultant DNA can be used as a template in a third round of mutagenesis, and so on.

A gene encoding (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase or myrcene synthase may be incorporated into any organism (intact plant, animal, microbe, etc.), or cell culture derived therefrom, that produces geranyl diphosphate. A (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase or myrcene synthase gene may be introduced into any organism for a variety of purposes including, but not limited to: production of (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase or myrcene synthase, or their products; production or modification of flavor and aroma properties; improvement of defense capability, and the alteration of other ecological interactions mediated by (-)-camphene, (-)-β-phellandrene, terpinolene, myrcene, (-)-limonene, (-)-pinene, or their derivatives.

Eukaryotic expression systems may be utilized for the production of (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase and myrcene synthase since they are capable of carrying out any required posttranslational modifications and of directing the enzymes to the proper membrane location. A representative eukaryotic expression system for this purpose uses the recombinant baculovirus, *Autographa californica* nuclear polyhedrosis virus (AcNPV; M.D. Summers and G.E. Smith, *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures* [1986]; Luckow et al., *Bio-technology* 6:47-55 [1987]) for expression of the terpenoid synthases of the invention. Infection of insect cells (such as cells of the species *Spodoptera frugiperda*) with the recombinant baculoviruses allows for the production of large amounts of the monoterpene synthase proteins. In addition, the baculovirus system has other important advantages for the production of recombinant monoterpene synthases. For example, baculoviruses do not infect humans and can therefore be safely handled in large quantities. In the baculovirus system, a DNA construct is prepared including a DNA segment encoding (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase or myrcene synthase and a vector. The vector may comprise the polyhedron gene promoter region of a baculovirus, the baculovirus flanking sequences necessary for proper cross-over during recombination (the flanking sequences comprise about 200-300 base pairs adjacent to the promoter sequence) and a bacterial origin of replication which permits the construct to replicate in bacteria. The vector is constructed so that (i) the DNA segment is placed adjacent (or operably linked or "downstream" or "under the control of") to the polyhedron gene promoter and (ii) the promoter/monoterpene synthase combination is flanked on both sides by 200-300 base pairs of baculovirus DNA (the flanking sequences).

To produce the monoterpene synthase DNA construct, a cDNA clone encoding the full length (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase or myrcene synthase is obtained using methods such as those described herein. The DNA construct is contacted in a host cell with baculovirus DNA of an appropriate baculovirus (that is, of the same species of baculovirus as the promoter encoded in the construct) under conditions such that recombination is effected. The resulting recombinant baculoviruses encode the full (-)-camphene

synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase or myrcene synthase. For example, an insect host cell can be cotransfected or transfected separately with the DNA construct and a functional baculovirus. Resulting recombinant baculoviruses can then be isolated and used to infect cells to effect production of the monoterpene synthase. Host insect cells include, for example, *Spodoptera frugiperda* cells, that are capable of producing a baculovirus-expressed monoterpene synthase. Insect host cells infected with a recombinant baculovirus of the present invention are then cultured under conditions allowing expression of the baculovirus-encoded (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase or myrcene synthase. Monoterpene synthases thus produced are then extracted from the cells using methods known in the art.

Other eukaryotic microbes such as yeasts may also be used to practice this invention. The baker's yeast *Saccharomyces cerevisiae*, is a commonly used yeast, although several other strains are available. The plasmid YRp7 (Stinchcomb et al., *Nature* **282**:39 [1979]; Kingsman et al., *Gene* **7**:141 [1979]; Tschemper et al., *Gene* **10**:157 [1980]) is commonly used as an expression vector in *Saccharomyces*. This plasmid contains the trp1 gene that provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, such as strains ATCC No. 44,076 and PEP4-1 (Jones, *Genetics* **85**:12 [1977]). The presence of the trp1 lesion as a characteristic of the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan. Yeast host cells are generally transformed using the polyethylene glycol method, as described by Hinnen (*Proc. Natl. Acad. Sci. USA* **75**:1929 [1978]). Additional yeast transformation protocols are set forth in Gietz et al., *N.A.R.* **20(17)** 1425(1992); Reeves et al., *FEMS* **99(2-3)**:193-197, (1992).

Suitable promoting sequences in yeast vectors include the promoters for 3-phosphoglycerate kinase (Hitzeman et al., *J. Biol. Chem.* **255**:2073 [1980]) or other glycolytic enzymes (Hess et al., *J. Adv. Enzyme Reg.* **7**:149 [1968]; Holland et al., *Biochemistry* **17**:4900 [1978]), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. In the construction of suitable expression plasmids, the termination sequences associated with these genes

are also ligated into the expression vector 3' of the sequence desired to be expressed to provide polyadenylation of the mRNA and termination. Other promoters that have the additional advantage of transcription controlled by growth conditions are the promoter region for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Any plasmid vector containing yeast-compatible promoter, origin of replication and termination sequences is suitable.

Cell cultures derived from multicellular organisms, such as plants, may be used as hosts to practice this invention. Transgenic plants can be obtained, for example, by transferring plasmids that encode (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase and/or myrcene synthase and a selectable marker gene, e.g., the kan gene encoding resistance to kanamycin, into *Agrobacterium tumifaciens* containing a helper Ti plasmid as described in Hoeckema et al., *Nature* **303:**179-181 [1983] and culturing the *Agrobacterium* cells with leaf slices of the plant to be transformed as described by An et al., *Plant Physiology* **81:**301-305 [1986]. Transformation of cultured plant host cells is normally accomplished through *Agrobacterium tumifaciens*, as described above. Cultures of mammalian host cells and other host cells that do not have rigid cell membrane barriers are usually transformed using the calcium phosphate method as originally described by Graham and Van der Eb (*Virology* **52:**546 [1978]) and modified as described in sections 16.32-16.37 of Sambrook et al., *supra.* However, other methods for introducing DNA into cells such as Polybrene (Kawai and Nishizawa, *Mol. Cell. Biol.* **4:**1172 [1984]), protoplast fusion (Schaffner, *Proc. Natl. Acad. Sci. USA* 77:2163 [1980]), electroporation (Neumann et al., *EMBO J.* **1:**841 [1982]), and direct microinjection into nuclei (Capecchi, *Cell* **22:**479 [1980]) may also be used. Additionally, animal transformation strategies are reviewed in Monastersky G.M. and Robl, J.M., *Strategies in Transgenic Animal Science*, ASM Press, Washington, D.C., 1995. Transformed plant calli may be selected through the selectable marker by growing the cells on a medium containing, e.g., kanamycin, and appropriate amounts of phytohormone such as naphthalene acetic acid and benzyladenine for callus and shoot induction. The plant cells may then be regenerated and the resulting plants transferred to soil using techniques well known to those skilled in the art.

In addition, a gene regulating (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase or myrcene synthase production can be incorporated into the plant along with a necessary promoter which is inducible. In the practice of this embodiment of the invention, a promoter that only responds to a specific external or internal stimulus is fused to the target cDNA. Thus, the gene will not be transcribed except in response to the specific stimulus. As long as the gene is not being transcribed, its gene product is not produced.

An illustrative example of a responsive promoter system that can be used in the practice of this invention is the glutathione-S-transferase (GST) system in maize. GSTs are a family of enzymes that can detoxify a number of hydrophobic electrophilic compounds that often are used as pre-emergent herbicides (Weigand et al., *Plant Molecular Biology* 7:235-243 [1986]). Studies have shown that the GSTs are directly involved in causing this enhanced herbicide tolerance. This action is primarily mediated through a specific 1.1 kb mRNA transcription product. In short, maize has a naturally occurring quiescent gene already present that can respond to external stimuli and that can be induced to produce a gene product. This gene has previously been identified and cloned. Thus, in one embodiment of this invention, the promoter is removed from the GST responsive gene and attached to a (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase or myrcene synthase gene that previously has had its native promoter removed. This engineered gene is the combination of a promoter that responds to an external chemical stimulus and a gene responsible for successful production of (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase or myrcene synthase.

In addition to the methods described above, several methods are known in the art for transferring cloned DNA into a wide variety of plant species, including gymnosperms, angiosperms, monocots and dicots (see, e.g., Glick and Thompson, eds., *Methods in Plant Molecular Biology*, CRC Press, Boca Raton, Florida [1993]). Representative examples include electroporation-facilitated DNA uptake by protoplasts (Rhodes et al., *Science* **240(4849):**204-207 [1988]); treatment of protoplasts with polyethylene glycol (Lyznik et al., *Plant Molecular Biology* **13:**151-161 [1989]); and bombardment of cells with DNA laden microprojectiles (Klein et al., *Plant Physiol.* **91:**440-444 [1989] and Boynton et al.,

Science **240(4858):**1534-1538 [1988]). Additionally, plant transformation strategies and techniques are reviewed in Birch, R.G., *Ann Rev Plant Phys Plant Mol Biol* **48:**297 (1997); Forester et al., *Exp. Agric.* **33:**15-33 (1997). Minor variations make these technologies applicable to a broad range of plant species.

Each of these techniques has advantages and disadvantages. In each of the techniques, DNA from a plasmid is genetically engineered such that it contains not only the gene of interest, but also selectable and screenable marker genes. A selectable marker gene is used to select only those cells that have integrated copies of the plasmid (the construction is such that the gene of interest and the selectable and screenable genes are transferred as a unit). The screenable gene provides another check for the successful culturing of only those cells carrying the genes of interest. A commonly used selectable marker gene is neomycin phosphotransferase II (NPT II). This gene conveys resistance to kanamycin, a compound that can be added directly to the growth media on which the cells grow. Plant cells are normally susceptible to kanamycin and, as a result, die. The presence of the NPT II gene overcomes the effects of the kanamycin and each cell with this gene remains viable. Another selectable marker gene which can be employed in the practice of this invention is the gene which confers resistance to the herbicide glufosinate (Basta). A screenable gene commonly used is the β-glucuronidase gene (GUS). The presence of this gene is characterized using a histochemical reaction in which a sample of putatively transformed cells is treated with a GUS assay solution. After an appropriate incubation, the cells containing the GUS gene turn blue.

The plasmid containing one or more of these genes is introduced into either plant protoplasts or callus cells by any of the previously mentioned techniques. If the marker gene is a selectable gene, only those cells that have incorporated the DNA package survive under selection with the appropriate phytotoxic agent. Once the appropriate cells are identified and propagated, plants are regenerated. Progeny from the transformed plants must be tested to insure that the DNA package has been successfully integrated into the plant genome.

Mammalian host cells may also be used in the practice of the invention. Examples of suitable mammalian cell lines include monkey kidney CVI line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line 293S (Graham et al., *J. Gen. Virol.* **36:**59 [1977]); baby hamster kidney cells (BHK, ATCC CCL 10); Chinese hamster ovary cells (Urlab and Chasin, *Proc. Natl. Acad. Sci USA* **77:**4216 [1980]); mouse sertoli cells (TM4, Mather, *Biol. Reprod.* **23:**243

[1980]); monkey kidney cells (CVI-76, ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1587); human cervical carcinoma cells (HELA, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor cells (MMT 060562, ATCC CCL 51); rat hepatoma cells (HTC, MI.54, Baumann et al., *J. Cell Biol.* **85:**1 [1980]); and TRI cells (Mather et al., *Annals N.Y. Acad. Sci.* **383:**44 [1982]). Expression vectors for these cells ordinarily include (if necessary) DNA sequences for an origin of replication, a promoter located in front of the gene to be expressed, a ribosome binding site, an RNA splice site, a polyadenylation site, and a transcription terminator site.

Promoters used in mammalian expression vectors are often of viral origin. These viral promoters are commonly derived from polyoma virus, Adenovirus 2, and most frequently Simian Virus 40 (SV40). The SV40 virus contains two promoters that are termed the early and late promoters. These promoters are particularly useful because they are both easily obtained from the virus as one DNA fragment that also contains the viral origin of replication (Fiers et al., *Nature* **273:**113 [1978]). Smaller or larger SV40 DNA fragments may also be used, provided they contain the approximately 250-bp sequence extending from the HindIII site toward the BglI site located in the viral origin of replication.

Alternatively, promoters that are naturally associated with the foreign gene (homologous promoters) may be used provided that they are compatible with the host cell line selected for transformation.

An origin of replication may be obtained from an exogenous source, such as SV40 or other virus (e.g., Polyoma, Adeno, VSV, BPV) and inserted into the cloning vector. Alternatively, the origin of replication may be provided by the host cell chromosomal replication mechanism. If the vector containing the foreign gene is integrated into the host cell chromosome, the latter is often sufficient.

The use of a secondary DNA coding sequence can enhance production levels of (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase and myrcene synthase in transformed cell lines. The secondary coding sequence typically comprises the enzyme dihydrofolate reductase (DHFR). The wild-type form of DHFR is normally inhibited by the chemical methotrexate (MTX). The level of DHFR expression in a cell will vary depending on the amount of MTX added to the cultured

host cells. An additional feature of DHFR that makes it particularly useful as a secondary sequence is that it can be used as a selection marker to identify transformed cells. Two forms of DHFR are available for use as secondary sequences, wild-type DHFR and MTX-resistant DHFR. The type of DHFR used in a particular host cell depends on whether the host cell is DHFR deficient (such that it either produces very low levels of DHFR endogenously, or it does not produce functional DHFR at all). DHFR-deficient cell lines such as the CHO cell line described by Urlaub and Chasin, *supra*, are transformed with wild-type DHFR coding sequences. After transformation, these DHFR-deficient cell lines express functional DHFR and are capable of growing in a culture medium lacking the nutrients hypoxanthine, glycine and thymidine. Nontransformed cells will not survive in this medium.

The MTX-resistant form of DHFR can be used as a means of selecting for transformed host cells in those host cells that endogenously produce normal amounts of functional DHFR that is MTX sensitive. The CHO-Kl cell line (ATCC No. CL 61) possesses these characteristics, and is thus a useful cell line for this purpose. The addition of MTX to the cell culture medium will permit only those cells transformed with the DNA encoding the MTX-resistant DHFR to grow. The nontransformed cells will be unable to survive in this medium.

Prokaryotes may also be used as host cells for the initial cloning steps of this invention. They are particularly useful for rapid production of large amounts of DNA, for production of single-stranded DNA templates used for site-directed mutagenesis, for screening many mutants simultaneously, and for DNA sequencing of the mutants generated. Suitable prokaryotic host cells include *E. coli* K12 strain 94 (ATCC No. 31,446), *E. coli* strain W3110 (ATCC No. 27,325) *E. coli* X1776 (ATCC No. 31,537), and *E. coli* B; however many other strains of *E. coli*, such as HB101, JM101, NM522, NM538, NM539, and many other species and genera of prokaryotes including bacilli such as *Bacillus subtilis*, other enterobacteriaceae such as *Salmonella typhimurium* or *Serratia marcesans*, and various *Pseudomonas* species may all be used as hosts. Prokaryotic host cells or other host cells with rigid cell walls are preferably transformed using the calcium chloride method as described in section 1.82 of Sambrook et al., *supra*. Alternatively, electroporation may be used for transformation of these cells. Prokaryote transformation techniques are set forth in Dower, W.J., in Genetic Engineering, Principles and Methods, **12**:275-296, Plenum Publishing Corp., 1990; Hanahan et al., *Meth. Enxymol.*, **204**:63 (1991).

As a representative example, cDNA sequences encoding (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase or myrcene synthase may be transferred to the $(His)_6$•Tag pET vector commercially available (from Novagen) for overexpression in *E. coli* as heterologous host. This pET expression plasmid has several advantages in high level heterologous expression systems. The desired cDNA insert is ligated in frame to plasmid vector sequences encoding six histidines followed by a highly specific protease recognition site (thrombin) that are joined to the amino terminus codon of the target protein. The histidine "block" of the expressed fusion protein promotes very tight binding to immobilized metal ions and permits rapid purification of the recombinant protein by immobilized metal ion affinity chromatography. The histidine leader sequence is then cleaved at the specific proteolysis site by treatment of the purified protein with thrombin, and the (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase and myrcene synthase again purified by immobilized metal ion affinity chromatography, this time using a shallower imidazole gradient to elute the recombinant synthases while leaving the histidine block still adsorbed. This overexpression-purification system has high capacity, excellent resolving power and is fast, and the chance of a contaminating *E. coli* protein exhibiting similar binding behavior (before and after thrombin proteolysis) is extremely small.

As will be apparent to those skilled in the art, any plasmid vectors containing replicon and control sequences that are derived from species compatible with the host cell may also be used in the practice of the invention. The vector usually has a replication site, marker genes that provide phenotypic selection in transformed cells, one or more promoters, and a polylinker region containing several restriction sites for insertion of foreign DNA. Plasmids typically used for transformation of *E. coli* include pBR322, pUC18, pUC19, pUCI18, pUC119, and Bluescript M13, all of which are described in sections 1.12-1.20 of Sambrook et al., *supra*. However, many other suitable vectors are available as well. These vectors contain genes coding for ampicillin and/or tetracycline resistance which enables cells transformed with these vectors to grow in the presence of these antibiotics.

The promoters most commonly used in prokaryotic vectors include the β-lactamase (penicillinase) and lactose promoter systems (Chang et al. *Nature* 375:615 [1978]; Itakura et al., *Science* 198:1056 [1977]; Goeddel et al.,

Nature **281**:544 [1979]) and a tryptophan (trp) promoter system (Goeddel et al., *Nucl. Acids Res.* **8**:4057 [1980]; EPO Appl. Publ. No. 36,776), and the alkaline phosphatase systems. While these are the most commonly used, other microbial promoters have been utilized, and details concerning their nucleotide sequences have been published, enabling a skilled worker to ligate them functionally into plasmid vectors (see Siebenlist et al., *Cell* **20**:269 [1980]).

Many eukaryotic proteins normally secreted from the cell contain an endogenous secretion signal sequence as part of the amino acid sequence. Thus, proteins normally found in the cytoplasm can be targeted for secretion by linking a signal sequence to the protein. This is readily accomplished by ligating DNA encoding a signal sequence to the 5' end of the DNA encoding the protein and then expressing this fusion protein in an appropriate host cell. The DNA encoding the signal sequence may be obtained as a restriction fragment from any gene encoding a protein with a signal sequence. Thus, prokaryotic, yeast, and eukaryotic signal sequences may be used herein, depending on the type of host cell utilized to practice the invention. The DNA and amino acid sequence encoding the signal sequence portion of several eukaryotic genes including, for example, human growth hormone, proinsulin, and proalbumin are known (see Stryer, *Biochemistry* W.H. Freeman and Company, New York, NY, p. 769 [1988]), and can be used as signal sequences in appropriate eukaryotic host cells. Yeast signal sequences, as for example acid phosphatase (Arima et al., *Nuc. Acids Res.* **11**:1657 [1983]), α–factor, alkaline phosphatase and invertase may be used to direct secretion from yeast host cells. Prokaryotic signal sequences from genes encoding, for example, LamB or OmpF (Wong et al., *Gene* **68**:193 [1988]), MalE, PhoA, or beta-lactamase, as well as other genes, may be used to target proteins from prokaryotic cells into the culture medium.

Trafficking sequences from plants, animals and microbes can be employed in the practice of the invention to direct the monoterpene synthase proteins of the present invention to the cytoplasm, endoplasmic reticulum, mitochondria or other cellular components, or to target the protein for export to the medium. These considerations apply to the overexpression of (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase and myrcene synthase, and to direction of expression within cells or intact organisms to permit gene product function in any desired location.

The construction of suitable vectors containing DNA encoding replication sequences, regulatory sequences, phenotypic selection genes and the monoterpene synthase DNA of interest are prepared using standard recombinant DNA procedures. Isolated plasmids and DNA fragments are cleaved, tailored, and ligated together in a specific order to generate the desired vectors, as is well known in the art (see, for example, Maniatis, *supra*, and Sambrook et al., *supra*).

As discussed above, (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase, (-)-limonene/(-)-α-pinene synthase, (-)-limonene synthase, (-)-pinene synthase and myrcene synthase variants are preferably produced by means of mutation(s) that are generated using the method of site-specific mutagenesis. This method requires the synthesis and use of specific oligonucleotides that encode both the sequence of the desired mutation and a sufficient number of adjacent nucleotides to allow the oligonucleotide to stably hybridize to the DNA template.

The foregoing may be more fully understood in connection with the following representative examples, in which "Plasmids" are designated by a lower case p followed by an alphanumeric designation. The starting plasmids used in this invention are either commercially available, publicly available on an unrestricted basis, or can be constructed from such available plasmids using published procedures. In addition, other equivalent plasmids are known in the art and will be apparent to the ordinary artisan.

"Digestion", "cutting" or "cleaving" of DNA refers to catalytic cleavage of the DNA with an enzyme that acts only at particular locations in the DNA. These enzymes are called restriction endonucleases, and the site along the DNA sequence where each enzyme cleaves is called a restriction site. The restriction enzymes used in this invention are commercially available and are used according to the instructions supplied by the manufacturers. (See also sections 1.60-1.61 and sections 3.38-3.39 of Sambrook et al., *supra*.)

"Recovery" or "isolation" of a given fragment of DNA from a restriction digest means separation of the resulting DNA fragment on a polyacrylamide or an agarose gel by electrophoresis, identification of the fragment of interest by comparison of its mobility versus that of marker DNA fragments of known molecular weight, removal of the gel section containing the desired fragment, and separation of the gel from DNA. This procedure is known generally. For example, see Lawn et al. (*Nucleic Acids Res.* 9:6103-6114 [1982]), and Goeddel et al. (*Nucleic Acids Res.*, *supra*).

The following examples merely illustrate the best mode now contemplated for practicing the invention, but should not be construed to limit the invention.

EXAMPLE 1

PCR-Based Generation of Probes for Cloning Certain Monoterpene Synthases From Grand fir (*Abies grandis*)

Substrates, Reagents and cDNA Library – [1-^{3}H]Geranyl diphosphate (250 Ci/mol) (Croteau, R., Alonso, W.R., Koepp, A.E., and Johnson, M.A. (1994) *Arch. Biochem. Biophys.* **309**:184-192), [1-^{3}H]farnesyl diphosphate (125 Ci/mol) (Dehal, S.S., and Croteau, R. (1988) *Arch. Biochem. Biophys.* **261**:346-356) and [1-^{3}H]geranylgeranyl diphosphate (120 Ci/mol) (LaFever, R.E., Stofer Vogel, B., and Croteau, R. (1994) *Arch. Biochem. Biophys.* **313**:139-149) were prepared as described previously. Terpenoid standards were from our own collection. All other biochemicals and reagents were purchased from Sigma Chemical Co. or Aldrich Chemical Co., unless otherwise noted. Construction of the λZAP II cDNA library, using mRNA isolated from wounded Grand fir sapling stems, was described previously (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268).

PCR-Based Probe Generation – Based on comparison of sequences of limonene synthase from spearmint (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024), 5-*epi*-aristolochene synthase from tobacco (Facchini, P.J., and Chappell, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**:11088-11092), and casbene synthase from castor bean (Mau, C.J.D., and West, C.A. (1994) *Proc. Natl. Acad. Sci. USA* **91**:8497-8501), four conserved regions were identified for which a set of consensus degenerate primers were synthesized: Primer A (SEQ ID NO:7); Primer B (SEQ ID NO:8); Primer C (SEQ ID NO:9); Primer D (SEQ ID NO:10). Primers A (SEQ ID NO:7), B (SEQ ID NO:8) and C (SEQ ID NO:9) have been described previously (Steele, C., Lewinsohn, E. and Croteau, R., *Proc. Nat'l. Acad. Sci. USA*, **92:** 4164-4168 (1995)); primer D (SEQ ID NO:10) was designed based on the conserved amino acid sequence motif DD(T/I)(I/Y/F)D(A/V)Y(A/G)(SEQ ID NO:25) of the above noted terpene synthases (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024; Facchini, P.J., and Chappell, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**:11088-11092; Mau, C.J.D., and West, C.A. (1994) *Proc. Natl. Acad. Sci. USA* **91**:8497-8501).

Each of the sense primers, A (SEQ ID NO:7), B (SEQ ID NO:8) and D (SEQ ID NO:10), was used for PCR in combination with antisense primer C (SEQ ID NO:9) by employing a broad range of amplification conditions. PCR was performed in a total volume of 50 µl containing 20 mM Tris/HCl (tris(hydroxymethyl) aminomethane/HCl, pH 8.4), 50 mM KCl, 5 mM $MgCl_2$, 200 µM of each dNTP, 1-5 µM of each primer, 2.5 units of *Taq* polymerase (BRL) and 5 µl of purified Grand fir stem cDNA library phage as template (1.5×10^9 pfu/ml). Analysis of the PCR reaction products by agarose gel electrophoresis (Sambrock, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) revealed that only the combination of primers C and D generated a specific PCR product of approximately 110 bps (base pairs). This PCR product was gel purified, ligated into pT7Blue (Novagen), and transformed into *E. coli* XL1-Blue cells. Plasmid DNA was prepared from 41 individual transformants and the inserts were sequenced (DyeDeoxy Terminator Cycle Sequencing, Applied Biosystems). Four different insert sequences were identified, and were designated as probes 1 (SEQ ID NO:11), 2 (SEQ ID NO:12), 4 (SEQ ID NO:13) and 5 (SEQ ID NO:14).

Subsequent isolation of four new cDNA species (*AG1.28* (SEQ ID NO:15); *AG2.2* (SEQ ID NO:1); *AG4.30* (SEQ ID NO:17) and *AG5.9* (SEQ ID NO:19)), encoding terpene synthases from Grand fir corresponding to probes 1 (SEQ ID NO:11), 2 (SEQ ID NO:12), 4 (SEQ ID NO:13) and 5 (SEQ ID NO:14), respectively, allowed the identification of three additional conserved sequence elements which were used to design a set of three new PCR primers: Primer E (5'-GGI GA(A/G) A(A/C)(A/G) (A/G)TI ATG GA(A/G) GA(A/G) GC-3')(SEQ ID NO:21); Primer F (5'-GA(A/G) (C/T)TI CA(G/A) (C/T)TI (A/C/T)(C/G/T)I (A/C)GI TGG TGG-3')(SEQ ID NO:22) and Primer G (5'-CCA (A/G)TT IA(A/G) ICC (C/T)TT IAC (A/G)TC-3')(SEQ ID NO:23).

Degenerate primer E (SEQ ID NO:21) was designed to conserved element GE(K/T)(V/I)M(E/D)EA (SEQ ID NO:26) and degenerate primer F (SEQ ID NO:22) was designed to conserved element Q(F/Y/D)(I/L)(T/L/R)RWW (SEQ ID NO:27) by comparing the sequences of five cloned terpene synthases from Grand fir: a monoterpene synthase corresponding to probe 2 (SEQ ID NO: 12), two sesquiterpene synthases corresponding to probe 4 (SEQ ID NO:13) and probe 5 (SEQ ID NO:14), respectively, a previously described diterpene synthase (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol.*

Chem. 271:23262-23268), and a truncated terpene synthase corresponding to probe 1 (SEQ ID NO:11). Degenerate primer G (SEQ ID NO:23) was designed according to the amino acid sequence DVIKG(F/L)NW (SEQ ID NO:28) obtained from a peptide generated by trypsin digestion of purified (-)-pinene synthase from Grand fir. Primers E (SEQ ID NO:21) and F (SEQ ID NO:22) were independently used for PCR amplification in combination with primer G (SEQ ID NO:23), with Grand fir stem cDNA library as template. The combination of primers E (SEQ ID NO:21) and G (SEQ ID NO:23) yielded a specific PCR product of approximately 1020 bps. This PCR product was ligated into pT7Blue and transformed into *E. coli* XL1-Blue. Plasmid DNA was prepared from 20 individual transformants and inserts were sequenced from both ends. The sequence of this 1022 bp insert was identical for all 20 plasmids and was designated as probe 3 (SEQ ID NO:24).

EXAMPLE 2

Screening a Wounded Grand fir Stem cDNA Library

For library screening, 100 ng of each probe was amplified by PCR, gel purified, randomly labeled with [α-^{32}P]dATP (Feinberg, A.P., and Vogelstein, B. (1984) *Anal. Biochem.* 137:266-267), and used individually to screen replica filters of 10^5 plaques of the wound-induced Grand fir stem cDNA library plated on *E. coli* LE392. Hybridization with probes 1 (SEQ ID NO:11), 2 (SEQ ID NO:12), 4 (SEQ ID NO:13) and 5 (SEQ ID NO:14) was performed for 14 h at 65 °C in 3 x SSPE and 0.1% SDS. Filters were washed three times for 10 min at 55 °C in 3 x SSPE with 0.1% SDS and exposed for 12 h to Kodak XAR film at -70 °C. All of the λZAPII clones yielding positive signals were purified through a second round of hybridization (probe 1 (SEQ ID NO:11) gave 25 positives, probe 2 (SEQ ID NO:12) gave 16 positives, probe 4 (SEQ ID NO:13) gave 49 positives and probe 5 (SEQ ID NO:14) gave 12 positives).

Hybridization with probe 3 (SEQ ID NO:24) was performed as before, but the filters were washed three times for 10 min at 65 °C in 3 x SSPE and 0.1% SDS before exposure. Approximately 400 λZAPII clones yielded strong positive signals, and 34 of these were purified through a second round of hybridization at 65°C. Approximately 400 additional clones yielded weak positive signals with probe 3 (SEQ ID NO:24), and 18 of these were purified through a second round of hybridization for 20 h at 45 °C. Purified λZAP II clones isolated using all five probes were *in vivo* excised as Bluescript II SK(-) phagemids and transformed into *E. coli* XLOLR according to the manufacturer's instructions (Stratagene). The size of each cDNA

insert was determined by PCR using T3 (SEQ ID NO:29) and T7 (SEQ ID NO:30) promoter primers and selected inserts (>1.5 kb) were partially sequenced from both ends.

EXAMPLE 3

Grand fir Monoterpene Synthase cDNA Expression in *E. coli* and Enzyme Assays

Except for cDNA clones AG3.18 (SEQ ID NO:3) and AG3.48 (SEQ ID NO:31), all of the partially sequenced inserts were either truncated at the 5'-end, or were out of frame, or bore premature stop codons upstream of the presumptive methionine start codon. For the purpose of functional expression, a 2023 bp insert fragment, extending from nucleotides 75 to 2097 of the sequence set forth in SEQ ID NO:1, and a 1911 bp insert fragment, extending from nucleotide 1 to nucleotide 1910 of the sequence set forth in SEQ ID NO:3, were subcloned in frame into pGEX vectors (Pharmacia). A 2016 bp fragment extending from nucleotide 73 to nucleotide 2088 of the sequence set forth in SEQ ID NO:5 was subcloned in frame into the pSBETa vector (Schenk, P.M., Baumann, S., Mattes, R., and Steinbiss, H.-H. (1995) *Biotechniques* 19, 196-200). To introduce suitable restriction sites for subcloning, fragments were amplified by PCR using primer combinations 2.2-*Bam*HI (5'-CAA AGG GAT CCA GAA TGG CTC TGG-3')(SEQ ID NO:33) and 2.2-*Not*I (5'-AGT AAG CGG CCG CTT TTT AAT CAT ACC CAC-3')(SEQ ID NO:34) with pAG2.2 insert (SEQ ID NO:1) as template, 3.18-*Eco*RI (5'-CTG CAG GAA TTC GGC ACG AGC-3')(SEQ ID NO:35) and 3.18-*Sma*I (5'-CAT AGC CCC GGG CAT AGA TTT GAG CTG-3')(SEQ ID NO:36) with pAG3.18 insert (SEQ ID NO:3) as template, and 10-*Nde*I (5-GGC AGG AAC ATA TGG CTC TCC TTT CTA TCG-3')(SEQ ID NO:37) and 10-*Bam*HI (5'-TCT AGA ACT AGT GGATCC CCC GGG CTG CAG-3')(SEQ ID NO:38) with pAG10 insert (SEQ ID NO:5) as template.

PCR reactions were performed in volumes of 50 µl containing 20 mM Tris/HCl (pH 8.8), 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100, 5 µg bovine serum albumin (BSA), 200 µM of each dNTP, 0.1 µM of each primer, 2.5 units of recombinant *Pfu* polymerase (Stratagene) and 100 ng plasmid DNA with the following program: denaturation at 94 °C, 1 min; annealing at 60 °C, 1 min; extension at 72 °C, 3.5 min; 35 cycles with final extension at 72 °C, 5 min. The PCR products were purified by agarose gel electrophoresis and used as template for a secondary PCR amplification with the identical conditions in total volumes of 250 µl each. Products from this secondary amplification were digested with the above

indicated restriction enzymes, purified by ultrafiltration and then ligated, respectively, into *Bam*HI/*Not*I-digested pGEX-4T-2 to yield plasmid pGAG2.2, into *Eco*RI/*Sma*I-digested pGEX-4T-3 to yield plasmid pGAG3.18, and into *Nde*I/*Bam*HI-digested pSBETa to yield plasmid pSBAG10; these plasmids were then transformed into *E. coli* XL1-Blue or *E. coli* BL21(DE3).

For expression, bacterial strains *E. coli* XLOLR/pAG3.18, *E. coli* XLOLR/pAG3.48, *E. coli* XL1-Blue/pGAG2.2, *E. coli* XL1-Blue/pGAG3.18, and *E. coli* BL21(DE3)/pSBAG10 were grown to A_{600} = 0.5 at 37 °C in 5 ml of LB medium (Sambrock, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) supplemented with 100 µg ampicillin/ml or 30 µg kanamycin/ml as determined by the vector. Cultures were then induced by addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside and grown for another 12 h at 20 °C. Cells were harvested by centrifugation (2000 x *g*, 10 min) and resuspended in either 1 ml monoterpene synthase assay buffer [50 mM Tris/HCl (pH 7.5), 500 mM KCl, 1 mM $MnCl_2$, 5 mM dithiothreitol, 0.05% (w/v) $NaHSO_3$ and 10% (v/v) glycerol], 1 ml sesquiterpene synthase assay buffer [10 mM dibasic potassium phosphate, 1.8 mM monobasic potassium phosphate (pH 7.3), 140 mM NaCl, 10 mM $MgCl_2$, 5 mM dithiothreitol, 0.05% (w/v) $NaHSO_3$ and 10% (v/v) glycerol], or 1 ml diterpene synthase assay buffer [30 mM Hepes (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid, pH 7.2), 7.5 mM $MgCl_2$, 5 mM dithiothreitol, 10 µM $MnCl_2$, 0.05% (w/v) $NaHSO_3$ and 10% (v/v) glycerol].

Cells were disrupted by sonication (Braun-Sonic 2000 with microprobe at maximum power for 15 seconds at 0-4°C), the homogenates were cleared by centrifugation (18,000 x *g*, 10 min), and 1 ml of the resulting supernatant was assayed for monoterpene synthase activity with 2.5 µM of [1-^{3}H]geranyl diphosphate, for sesquiterpene synthase activity with 3.5 µM [1-^{3}H]farnesyl diphosphate, or for diterpene synthase activity with 5 µM [1-^{3}H]geranylgeranyl diphosphate following standard protocols (Croteau, R., and Cane, D.E. (1985) *Methods Enzymol.* **110**:383-405; LaFever, R.E., Stofer Vogel, B., and Croteau, R. (1994) *Arch. Biochem. Biophys.* **313**:139-149; Dehal, S.S., and Croteau, R. (1988) *Arch. Biochem. Biophys.* **261**:346-356). In the case of the monoterpene synthase and sesquiterpene synthase assays, the incubation mixture was overlaid with 1 ml pentane to trap volatile products. In all cases, after incubation at 31°C for 2 h, the reaction mixture was extracted with pentane (3 x 1 ml) and the combined extract was passed through a 1.5

ml column of anhydrous $MgSO_4$ and silica gel (Mallinckrodt 60 Å) to provide the terpene hydrocarbon fraction free of oxygenated metabolites. The columns were subsequently eluted with 3 x 1 ml of ether to collect any oxygenated products, and an aliquot of each fraction was taken for liquid scintillation counting to determine conversion rate.

Product Identification – To obtain sufficient product for analysis by radio-GLC (gas liquid chromatography), chiral capillary GLC and GLC-MS (mass spectrum/spectrometry), preparative-scale enzyme incubations were carried out. Thus, the enzyme was prepared from 50 ml of cultured bacterial cells by extraction with 3 ml of assay buffer as above, and the extracts were incubated with excess substrate overnight at 31°C. The hydrocarbon fraction was isolated by elution through $MgSO_4$-silica gel as before, and the pentane eluate was concentrated for evaluation by capillary radio-GLC as described (Croteau, R., and Satterwhite, D.M. (1990) *J. Chromatogr.* **500**:349-354), by chiral column capillary GLC (Lewinsohn, E., Savage, T.J., Gijzen, M., and Croteau, R. (1993) *Phytochem. Anal.* 4:220-225), and by combined GLC-MS [Hewlett-Packard 6890 GC-MSD with cool (40°C) on-column injection, detection via electron impact ionization (70 eV), He carrier at 0.7 psi., column: 0.25 mm i.d. x 30 m fused silica with 0.25 µm film of 5MS (Hewlett-Packard) programmed from 35°C (5 min hold) to 230°C at 5°C/min].

EXAMPLE 4

Sequence Analysis

Inserts of all recombinant bluescript plasmids and pGEX plasmids were completely sequenced on both strands via primer walking and nested deletions (Sambrock, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) using the DyeDeoxy Terminator Cycle Sequencing method (Applied Biosystems). Sequence analysis was done using the Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI.

EXAMPLE 5

RNA Extraction and Northern Blotting

Grand fir sapling stem tissue was harvested prior to wounding or two days after wounding by a standard procedure (Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**:267-273). Total RNA was isolated (Lewinsohn, E., Steele, C.L., and Croteau, R. (1994) *Plant Mol. Biol. Rep.* **12**:20-25) and 20 µg of RNA per gel lane was separated under denaturing conditions (Sambrock, J.,

Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) and transferred to nitrocellulose membranes (Schleicher and Schuell) according to the manufacturer's protocol. To prepare hybridization probes, cDNA fragments of 1.4-1.5 kb were amplified by PCR from *AG2.2* (SEQ ID NO:1) with primer JB29 (5'-CTA CCA TTC CAA TAT CTG-3')(SEQ ID NO:39) and primer 2-8 (5'-GTT GGA TCT TAG AAG TTC CC-3')(SEQ ID NO:40), from *AG3.18* (SEQ ID NO:3) with primer 3-9 (5'-TTT CCA TTC CAA CCT CTG GG-3')(SEQ ID NO:41) and primer 3-11 (5'-CGT AAT GGA AAG CTC TGG CG-3')(SEQ ID NO:42), and from *AG10* (SEQ ID NO:5) with primer 7-1 (5'-CCT TAC ACG CCT TTG GAT GG-3')(SEQ ID NO:43) and primer 7-3 (5'-TCT GTT GAT CCA GGA TGG TC-3')(SEQ ID NO:44). The probes were randomly labeled with [α-^{32}P]dATP (Feinberg, A.P., and Vogelstein, B. (1984) *Anal. Biochem.* **137**:266-267). Blots were hybridized for 24 h at 55°C in 3 x SSPE and 0.1% SDS, washed at 55°C in 1 x SSPE and 0.1%, SDS and subjected to autoradiography as described above at -80°C for 24 h.

EXAMPLE 6

Cloning and Characterization of Clones AG1.28 (SEQ ID NO:15), AG2.2 (SEQ ID NO:1), AG4.30 (SEQ ID NO:17) and AG 5.9 (SEQ ID NO:19)

Similarity-Based Cloning of Grand fir Terpene Synthases – Grand fir has been developed as a model system for the study of induced oleoresin production in conifers in response to wounding and insect attack (Johnson, M.A., and Croteau, R. (1987) in *Ecology and Metabolism of Plant Lipids* (Fuller, G., and Nes, W.D., eds) pp. 76-91, American Chemical Society Symposium Series 325, Washington, DC; Gijzen, M., Lewinsohn, E., Savage, T.J., and Croteau, R.B. (1993) in *Bioactive Volatile Compounds from Plants* (Teranishi, R., Buttery, R.G., and Sugisawa, H., eds) pp. 8-22, American Chemical Society Symposium Series 525, Washington, DC; Raffa, K. F., and Berryman, A.A. (1982) *Can. Entomol.* **114**:797-810; Steele, C., Lewinsohn, E., and Croteau, R. (1995) *Proc. Natl. Acad. Sci. USA* **92**:4164-4168; Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) in *Regulation of Isopentenoid Metabolism* (Nes, W.D., Parish, E.J., and Trzaskos, J.M., eds) pp. 8-17, American Chemical Society Symposium Series 497, Washington, DC). The chemistry and biosynthesis of the oleoresin monoterpenes, sesquiterpenes and diterpenes have been well defined (Lewinsohn, E., Savage, T.J., Gijzen, M., and Croteau, R. (1993) *Phytochem. Anal.* **4**:220-225; Lewinsohn, E., Gijzen, M., and Croteau, R. (1991) *Plant Physiol.* **96**:44-49; Funk, C., Lewinsohn, E., Stofer Vogel, B., Steele C., and

Croteau, R. (1994) *Plant Physiol.* **106**:999-1005; Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**:267-273; Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173; LaFever, R.E., Stofer Vogel, B., and Croteau, R. (1994) *Arch. Biochem. Biophys.* **313**:139-149; Funk, C., and Croteau, R. (1994) *Arch. Biochem. Biophys.* **308**:258-266); however, structural analysis of the responsible terpene synthases as well as studies on the regulation of oleoresinosis require the isolation of cDNA species encoding the terpene synthases. Protein purification from conifers, as the basis for cDNA isolation, has been of limited success (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268) and thus far has not permitted cloning of any of the monoterpene synthases from these species (Steele, C., Lewinsohn, E., and Croteau, R. (1995) *Proc. Natl. Acad. Sci. USA* **92**:4164-4168).

As a possible alternative to protein-based cloning of terpene synthases, a homology-based PCR strategy was proposed (Steele, C., Lewinsohn, E., and Croteau, R. (1995) *Proc. Natl. Acad. Sci. USA* **92**:4164-4168) that was founded upon the three terpene synthases of plant origin then available, a monoterpene synthase, (-)-4*S*-limonene synthase, from spearmint (*Mentha spicata*, Lamiaceae) (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024), a sesquiterpene synthase, 5-*epi*-aristolochene synthase, from tobacco (*Nicotiana tabacum*, Solanaceae) (Facchini, P.J., and Chappell, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**:11088-11092), and a diterpene synthase, casbene synthase, from castor bean (*Ricinus communis*, Euphorbiaceae) (Mau, C.J.D., and West, C.A. (1994) *Proc. Natl. Acad. Sci. USA* **91**:8497-8501). Despite the taxonomic distances between these three angiosperm species and the differences in substrate utilized, reaction mechanism and product type of the three enzymes, a comparison of the deduced amino acid sequences identified several conserved regions that appeared to be useful for the design of degenerate PCR primers (see Example 1). Using cDNA from a wound-induced Grand fir stem library as template, PCR primers C (SEQ ID NO:9) and D (SEQ ID NO:10) amplified products corresponding to four distinct sequence groups, all of which showed significant similarity to sequences of cloned terpene synthases of plant origin. The four different inserts were designated as probes 1 (SEQ ID NO:11), 2 (SEQ ID NO:12), 4 (SEQ ID NO:13) and 5 (SEQ ID NO:14), and were employed for isolation of the corresponding cDNA clones by plaque hybridization.

Screening of 10^5 cDNA phage plaques from the wounded Grand fir stem library, with each of the four probes, yielded a four-fold difference in the number of positives, most likely reflecting different levels of expression of the corresponding genes. Size selected inserts (>1.5 kb) of purified and *in vivo* excised clones were partially sequenced from both ends, and were shown to segregate into four distinct groups corresponding to the four hybridization probes. Since all cDNAs corresponding to probes 1 (SEQ ID NO:11), 4 (SEQ ID NO:13) and 5 (SEQ ID NO:14) were truncated at their 5'-ends, only inserts of the largest representatives of each group, clone *AG1.28* (SEQ ID NO:15), clone *AG2.2* (SEQ ID NO:1) (apparently full length), clone *AG4.30* (SEQ ID NO:17) and clone *AG5.9* (SEQ ID NO:19), were completely sequenced. Clone *AG1.28* (SEQ ID NO:15)(2424 bps) includes an open reading frame (ORF) of 2350 nucleotides (nts) encoding 782 amino acids (SEQ ID NO:16); clone *AG2.2* (SEQ ID NO:1)(2196 bps), includes an ORF of 1881 nts encoding 627 amino acids (SEQ ID NO:2); clone *AG4.30* (SEQ ID NO:17)(1967 bps) includes an ORF of 1731 nts encoding 577 amino acids (SEQ ID NO:18) and clone *AG5.9* (SEQ ID NO:19)(1416 bps) includes an ORF of 1194 nucleotides encoding 398 amino acids (SEQ ID NO:20).

cDNA clones *AG1.28* (SEQ ID NO:15), *AG2.2* (SEQ ID NO:1), *AG4.30* (SEQ ID NO:17) and *AG5.9* (SEQ ID NO:19) were compared pairwise with each other and with other cloned plant terpene synthases. Truncated clone *AG1.28* (SEQ ID NO:15) resembled most closely in size and sequence (72% similarity, 49% identity) a diterpene cyclase, abietadiene synthase, from Grand fir (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* 271:23262-23268). Clones *AG4.30* (SEQ ID NO:17) and *AG5.9* (SEQ ID NO:19) share approximately 80% similarity (60% identity) at the amino acid level, and are almost equally distant from both clone *AG1.28* (SEQ ID NO:15) and full-length clone *AG2.2* (SEQ ID NO:1)(range of 65-70% similarity and 45-47% identity); the amino acid sequence similarity between *AG1.28* (SEQ ID NO:15) and *AG2.2* (SEQ ID NO:1) is 65% (41% identity). Considering the high level of homology between *AG4.30* (SEQ ID NO:17) and *AG5.9* (SEQ ID NO:19), these comparisons suggest that the four new cDNAs, *AG1.28* (SEQ ID NO:15), *AG2.2* (SEQ ID NO:1), *AG4.30* (SEQ ID NO:17) and *AG5.9* (SEQ ID NO:19), represent the three major subfamilies of Grand fir terpene synthase genes encoding monoterpene synthases, sesquiterpene synthases and diterpene synthases.

Identification of cDNA Clone AG2.2 (SEQ ID NO:1) *as Myrcene Synthase* – The pAG2.2 insert (SEQ ID NO:1) appeared to be a full-length clone encoding a protein of molecular weight 72,478 with a calculated pI at 6.5. The size of the translated protein encoded by *AG2.2* (SEQ ID NO:1) (627 residues)(SEQ ID NO:2) is in the range of the monoterpene synthase preproteins for limonene synthase from spearmint (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024) and *Perilla frutescens* (Yuba, A., Yazaki, K., Tabata, M., Honda, G., and Croteau, R. (1996) *Arch. Biochem. Biophys.* **332**:280-287), but is about 240 amino acids shorter than the two gymnosperm diterpene synthase preproteins for abietadiene synthase (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268) and taxadiene synthase (Wildung, M.R., and Croteau, R. (1996) *J. Biol. Chem.* **271**:9201-9204). Monoterpene and diterpene biosynthesis are compartmentalized in plastids whereas sesquiterpene biosynthesis is cytosolic (reviewed in Kleinig, H. (1989) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**:39-53; Gershenzon, J., and Croteau, R. (1993) in *Lipid Metabolism in Plants* (Moore, T.S. Jr., ed) pp. 339-388, CRC Press, Boca Raton, FL; McGarvey, D.J., and Croteau, R. (1995) *Plant Cell* 7, 1015-1026); thus, monoterpene and diterpene synthases are encoded as preproteins bearing an amino-terminal transit peptide for import of these nuclear gene products into plastids where they are proteolytically processed to the mature forms (Keegstra, K., Olsen, J.J., and Theg, S.M. (1989) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**:471-501; von Heijne, G., Stepphuhn, J., and Herrmann (1989) *Eur. J. Biochem.* **180**:535-545). Both the size of the deduced protein and the presence of an *N*-terminal domain (of 60 to 70 amino acids) with features characteristic of a targeting sequence [rich in serine residues (16-18%) and low in acidic residues (four Asp or Glu) (Keegstra, K., Olsen, J.J., and Theg, S.M. (1989) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**:471-501; von Heijne, G., Stepphuhn, J., and Herrmann (1989) *Eur. J. Biochem.* **180**:535-545)] suggest that *AG2.2* (SEQ ID NO:1) encodes a monoterpene synthase rather than a sesquiterpene synthase or a diterpene synthase.

Since pAG2.2 contained the terpene synthase insert in reversed orientation, the ORF was subcloned in frame with glutathione *S*-transferase, for ultimate ease of purification (Bohlmann, J., DeLuca, V., Eilert, U., and Martin, W. (1995) *Plant J.* 7:491-501; Bohlmann, J., Lins, T., Martin, W., and Eilert, U. (1996) *Plant Physiol.* **111**:507-514), into pGEX-4T-2, yielding plasmid pGAG2.2. The recombinant fusion

protein was expressed in *E. coli* strain XL1-Blue/pGAG2.2, then extracted and assayed for monoterpene synthase, sesquiterpene synthase and diterpene synthase activity using tritium labeled geranyl diphosphate, farnesyl diphosphate and geranylgeranyl diphosphate as the respective substrate. Enzymatic production of a terpene olefin was observed only with geranyl diphosphate as substrate, and the only product was shown to be myrcene by radio-GLC and GLC-MS comparison to an authentic standard (FIGURE 3). Bacteria transformed with pGEX vector containing the *AG2.2* insert (SEQ ID NO:1) in antisense orientation did not afford detectable myrcene synthase activity when induced, and the protein isolated and assayed as above. A myrcene synthase cDNA has not been obtained previously from any source, although myrcene is a minor co-product (2%) of the native and recombinant limonene synthase from spearmint (Rajaonarivony, J.I.M., Gershenzon, J., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **296**:49-57; Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024) and of several enzymes from sage (Croteau, R., and Satterwhite, D.M. (1989) *J. Biol. Chem.* **264**:15309-15315). cDNA cloning and functional expression of myrcene synthase, which is one of several wound-inducible monoterpene synthase activities of Grand fir (Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**:267-273), demonstrates that this acyclic monoterpene is formed by a distinct enzyme and is not a co-product of another synthase.

EXAMPLE 7

Cloning and Characterization of Clones AG3.18 (SEQ ID NO:3) Encoding (-)-Pinene Synthase and cDNA Clone AG10 (SEQ ID NO:5) Encoding (-)-Limonene Synthase

Identification of cDNA Clone AG3.18 (SEQ ID NO:3) *as (-)-Pinene Synthase and cDNA Clone AG10* (SEQ ID NO:5) *as (-)-Limonene Synthase* – Alignment of the four new terpene synthase cDNA sequences (*AG1.28* (SEQ ID NO:15), *AG2.2* (SEQ ID NO:1), *AG4.30* (SEQ ID NO:17) and *AG.5.9* (SEQ ID NO:19)), and that for abietadiene synthase (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268), allowed the identification of several conserved sequence motifs among this enzyme family from Grand fir, which provided the foundation for an extended similarity-based cloning approach. Two new sense primers E (SEQ ID NO:21) and F (SEQ ID NO:22) were designed according to conserved sequence elements, whereas a degenerate antisense primer G (SEQ ID NO:23) was designed based upon very limited amino acid sequence information from

pinene synthase (see Example 1). Only the combination of primers E (SEQ ID NO:21) and G (SEQ ID NO:23) amplified a specific product of 533 bps, which was designated as probe 3 (SEQ ID NO:24).

Hybridization of 10^5 Grand fir λZAP II cDNA clones with probe 3 (SEQ ID NO:24) yielded two types of signals comprised of about 400 strongly positive clones and an equal number of weak positives, indicating that the probe recognized more than one type of cDNA. Thirty-four of the former clones and 18 of the latter were purified, the inserts were selected by size (2.0-2.5 kb), and the *in vivo* excised clones were partially sequenced from both ends. Those clones which afforded weak hybridization signals were shown to contain inserts that were either identical to myrcene synthase clone *AG2.2* (SEQ ID NO:1) or exhibited no significant sequence similarity to terpene synthases. Clone AG3.48 (SEQ ID NO:31) contained the myrcene synthase ORF in the correct orientation and in frame for expression from the Bluescript plasmid vector. This cDNA was functionally expressed in *E. coli* and the resulting enzyme was shown to accept only geranyl diphosphate as the prenyl diphosphate substrate and to produce myrcene as the exclusive reaction product. This finding with AG3.48 (SEQ ID NO:31) confirms that expression of AG2.2 (SEQ ID NO:1) as the glutathione *S*-transferase fusion protein from pGAG2.2 does not influence substrate utilization or product outcome of the myrcene synthase.

Clones which gave strong hybridization signals segregated into distinct sequence groups represented by clone *AG3.18* (SEQ ID NO:3)(2018 bp insert with ORF of 1884 nt; encoded protein of 628 residues at 71,505 Da and pI of 5.5) and *AG10* (SEQ ID NO:5)(2089 bp insert with ORF of 1911 nt; encoded protein of 637 residues at 73,477 Da and pI of 6.4). *AG3.18* (SEQ ID NO:3) and *AG10* (SEQ ID NO:5) form a subfamily together with the myrcene synthase clone *AG2.2* (SEQ ID NO:1) that is characterized by a minimum of 79% pairwise similarity (64% identity) at the amino acid level. Like myrcene synthase, both *AG3.18* (SEQ ID NO:3) and *AG10* (SEQ ID NO:5) encode *N*-terminal sequences of 60 to 70 amino acids which are rich in serine (19-22% and 11-15%, respectively) and low in acidic residues (4 and 2, respectively) characteristic of plastid transit peptides (Keegstra, K., Olsen, J.J., and Theg, S.M. (1989) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**:471-501; von Heijne, G., Stepphuhn, J., and Herrmann (1989) *Eur. J. Biochem.* **180**:535-545).

Plasmid pAG3.18 (SEQ ID NO:3) contained the presumptive terpene synthase ORF in frame for direct expression from the bluescript plasmid, whereas the *AG10* (SEQ ID NO:5) ORF was in reversed orientation. Both *AG3.18* (SEQ ID NO:3) and

AG10 (SEQ ID NO:5) were subcloned into expression vectors yielding plasmids pGAG3.18 and pSBAG10. Recombinant proteins were expressed in bacterial strain *E. coli* XLOLR/pAG3.18, *E. coli* XL1-Blue/pGAG3.18 and *E. coli* BL21(DE3)/pSBAG10. When extracts of the induced cells were tested for terpene synthase activity with all of the potential prenyl diphosphate substrates, only geranyl diphosphate was utilized. Extracts from *E. coli* BL21(DE3)/pSBAG10 converted geranyl diphosphate to limonene as the major product with lesser amounts of α-pinene, β-pinene and β-phellandrene, as determined by radio-GLC and combined GLC-MS (FIGURE 5). Chiral phase capillary GLC on β-cyclodextrin revealed the limonene product to be the (-)-4*S*-enantiomer and the pinene products to be the related (-)-(1*S*:5*S*)-enantiomers. Although optically pure standards were not available for the analysis, stereochemical considerations suggest that the minor product β-phellandrene is also the mechanistically related (-)-(4*S*)-antipode (Gambliel, H., and Croteau, R. (1984) *J. Biol. Chem.* **259**:740-748; Croteau, R., Satterwhite, D.M., Cane, D.E., and Chang, C.C. (1988) *J. Biol. Chem.* **263**:10063-10071; Wagschal, K., Savage, T.J., and Croteau, R. (1991) *Tetrahedron* **47**:5933-5944; Croteau, R., Satterwhite, D.M., Wheeler, C.J., and Felton, N.M. (1989) *J. Biol. Chem.* **264**:2075-2080; LaFever, R.E., and Croteau, R. (1993) *Arch. Biochem. Biophys.* **301**:361-366). Similar analysis of the monoterpene products generated from geranyl diphosphate by cell-free extracts of *E. coli* XLOLR/pAG3.18 and *E. coli* XL1-Blue/pGAG3.18 demonstrated the presence of a 42:58% mixture of α-pinene and β-pinene (FIGURE 4), the same product ratio previously described for the purified, native (-)-pinene synthase from Grand fir (Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173). Chiral phase capillary GLC confirmed the products of the recombinant pinene synthase to be the (-)-(1*S*:5*S*)-enantiomers, as expected. No other monoterpene co-products were detected with the recombinant (-)-pinene synthase, as observed previously for the native enzyme (Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173).

Evidence for the formation of both α- and β-pinene by a single enzyme has been previously provided through co-purification studies, and differential inhibition and inactivation studies, as well as by isotopically sensitive branching experiments (Gambliel, H., and Croteau, R. (1984) *J. Biol. Chem.* **259**:740-748; Wagschal, K.C., Pyun, H.-J., Coates, R.M., and Croteau, R. (1994) *Arch. Biochem. Biophys.* **308**:477-487; Wagschal, K., Savage, T.J., and Croteau, R. (1991) *Tetrahedron* **47**:5933-

5944; Croteau, R., Wheeler, C.J., Cane, D.E., Ebert, R., and Ha, H.-J. (1987) *Biochemistry* **26**:5383-5389). The cDNA cloning of pinene synthase provides the ultimate proof that a single enzyme forms both products. The calculated molecular weight of the (-)-pinene synthase deduced from *AG3.18* (SEQ ID NO:3) is approximately 64,000 (excluding the putative transit peptide), which agrees well with the molecular weight of 63,000 established for the native enzyme from Grand fir by gel permeation chromatography and SDS-PAGE (Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173).

A limonene synthase cDNA has thus far been cloned only from two very closely related angiosperm species (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024; Yuba, A., Yazaki, K., Tabata, M., Honda, G., and Croteau, R. (1996) *Arch. Biochem. Biophys.* **332**:280-287), and the isolation of a pinene synthase cDNA has not been reported before. Pinene synthase has previously received considerable attention as a major defense-related monoterpene synthase in conifers (Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**:267-273; Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173). In the Grand fir cDNA library, which was synthesized from mRNA obtained from wound-induced sapling stems, clones corresponding to pinene synthase are at least ten times more abundant than clones for myrcene synthase. This finding reflects the relative proportions of the induced levels of activities of these enzymes in Grand fir saplings; pinene synthase and limonene synthase are the major monoterpene synthase activities whereas the induced level of myrcene synthase activity is relatively low (Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**:267-273). The cDNAs for inducible monoterpene synthases provide probes for genetic and molecular analysis of oleoresin-based defense in conifers. Northern blots (FIGURE 6) of total RNA extracted from non-wounded sapling stems and from stems two days after wounding (when enzyme activity first appears) were probed with cDNA fragments for *AG2.2* (SEQ ID NO:1), *AG3.18* (SEQ ID NO:3) and *AG10* (SEQ ID NO:5), and thus demonstrated that increased mRNA accumulation for monoterpene synthases is responsible for this induced, defensive response in Grand fir. The availability of cloned, defense-related monoterpene synthases presents several possible avenues for transgenic manipulation of oleoresin composition to improve tree resistance to bark beetles and other pests. For example, altering the monoterpene content of oleoresin may chemically disguise the host and decrease insect aggregation by changing the

levels of pheromone precursors or predator attractants, or lower infestation by increasing toxicity toward beetles and their pathogenic fungal associates (Johnson, M.A., and Croteau, R. (1987) in *Ecology and Metabolism of Plant Lipids* (Fuller, G., and Nes, W.D., eds) pp. 76-91, American Chemical Society Symposium Series 325, Washington, DC; Gijzen, M., Lewinsohn, E., Savage, T.J., and Croteau, R.B. (1993) in *Bioactive Volatile Compounds from Plants* (Teranishi, R., Buttery, R.G., and Sugisawa, H., eds) pp. 8-22, American Chemical Society Symposium Series 525, Washington, DC; Byers, J.A. (1995) in *Chemical Ecology of Insects 2* (Cardé, R.T., and Bell, W.J., eds) pp. 154-213, Chapman and Hall, New York).

EXAMPLE 8

Properties of the Recombinant Monoterpene Synthases Encoded by cDNA Clones *AG2.2 (SEQ ID NO:1), AG3.18 (SEQ ID NO:3)* and *AG10 (SEQ ID NO:5)*

All three recombinant enzymes require Mn^{2+} for activity, and Mg^{2+} is essentially ineffective as the divalent metal ion cofactor. This finding confirms earlier results obtained with the native monoterpene synthases of Grand fir and lodgepole pine (*Pinus contorta*) (Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173; Savage, T.J., Hatch, M.W., and Croteau, R. (1994) *J. Biol. Chem.* **269**:4012-4020). All terpene synthases and prenyltransferases are thought to employ a divalent metal ion, usually Mg^{2+} or Mn^{2+}, in the ionization steps of the reaction sequence to neutralize the negative charge of the diphosphate leaving group (Croteau, R. (1987) *Chem. Rev.* **87**:929-954; Cane, D.E. (1992) *Ciba Found. Symp. Ser.* **171**:163-167; Poulter, C.D., and Rilling, H.C. (1981) in *Biosynthesis of Isoprenoid Compounds* (Porter, J.W., and Spurgeon, S.L., eds) Vol. 1, pp. 161-224, Wiley & Sons, New York), and all relevant sequences thus far obtained bear a conserved aspartate rich element (DDXXD)(SEQ ID NO:45) considered to be involved in divalent metal ion binding (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268; Ashby, M.N., and Edwards, P.A. (1990) *J. Biol. Chem.* **265**:13157-13164; Chen, A., Kroon, P.A., and Poulter, D.C. (1994) *Protein Sci.* **3**:600-607; Tarshis, L.C., Yan, M., Poulter, C.D., and Sacchettini, J.C. (1994) *Biochemistry* **33**:10871-10877; Cane, D.E., Sohng, J.K., Lamberson, C.R., Rudnicki, S.M., Wu, Z., Lloyd, M.D., Oliver, J.S., and Hubbard, B.R. (1994) *Biochemistry* **33**:5846-5857; Proctor, R.H., and Hohn, T.M. (1993) *J. Biol. Chem.* **268**:4543-4548). In addition to this strict, general dependence on a divalent metal ion, the monoterpene synthases of conifers are unique in their further

requirement for a monovalent cation (K^+), a feature that distinguishes the gymnosperm monoterpene synthases from their counterparts from angiosperm species and implies a fundamental structural and/or mechanistic difference between these two families of catalysts (Savage, T.J., Hatch, M.W., and Croteau, R. (1994) *J. Biol. Chem.* **269**:4012-4020). All three recombinant monoterpene synthases depend upon K^+, with maximum activity achieved at approximately 500 mM KCl. A requirement for K^+ has been reported for a number of different types of enzymes, including those that catalyze phosphoryl cleavage or transfer reactions (Suelter, C.H. (1970) *Science* **168**:789-794) such as Hsc70 ATPase (Wilbanks, S.M., and McKay, D.B. (1995) *J. Biol. Chem.* **270**:2251-2257). The crystal structure of bovine Hsc70 ATPase indicates that both Mg^{2+} and K^+ interact directly with phosphate groups of the substrate and implicates three active site aspartate residues in Mg^{2+} and K^+ binding (Wilbanks, S.M., and McKay, D.B. (1995) *J. Biol. Chem.* **270**:2251-2257), reminiscent of the proposed role of the conserved DDXXD (SEQ ID NO:45) motif of the terpene synthases and prenyltransferases in divalent cation binding, a function also supported by recent site directed mutagenesis (Marrero, P.F., Poulter, C.D., and Edwards, P.A. (1992) *J. Biol. Chem.* **267**:21873-21878; Joly, A., and Edwards, P.A. (1993) *J. Biol. Chem.* **268**: 26983-26989; Song, L., and Poulter, C.D. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**:3044-3048; Koyama, T., Tajima, M., Sano, H., Doi, T., Koike-Takeshita, A., Obata, S., Nishino, T., and Ogura, K. (1996) *Biochemistry* **35**:9533-9538) and by X-ray structural analysis (Tarshis, L.C., Yan, M., Poulter, C.D., and Sacchettini, J.C. (1994) *Biochemistry* **33**:10871-10877) of farnesyl diphosphate synthase.

cDNA cloning and functional expression of the myrcene, limonene and pinene synthases from Grand fir represent the first example of the isolation of multiple synthase genes from the same species, and provide tools for evaluation of structure-function relationships in the construction of acyclic, monocyclic and bicyclic monoterpene products and for detailed comparison to catalysts from phylogenetically distant plants that carry out ostensibly identical reactions (Gambliel, H., and Croteau, R. (1984) *J. Biol. Chem.* **259**:740-748; Rajaonarivony, J.I.M., Gershenzon, J., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **296**:49-57; Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024; Adam, K.-P., Crock, J., and Croteau, R. (1996) *Arch. Biochem. Biophys.* **332**:352-356). The recent acquisition of cDNA isolates encoding sesquiterpene synthases and diterpene synthases (Stofer Vogel, B., Wildung, M.R.,

Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268) from Grand fir should, together with the monoterpene synthases, also permit addressing the structural basis of chain-length specificity for prenyl diphosphate substrates in this family of related enzymes.

EXAMPLE 9

Sequence Comparison of Certain Cloned Monoterpene Synthases

Previous studies based on substrate protection from inactivation with selective amino acid modifying reagents have implicated functionally important cysteine, histidine and arginine residues in a range of different monoterpene synthases (Rajaonarivony, J.I.M., Gershenzon, J., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **296**:49-57; Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173; Savage, T.J., Hatch, M.W., and Croteau, R. (1994) *J. Biol. Chem.* **269**:4012-4020; Rajaonarivony, J.I.M., Gershenzon, J., Miyazaki, J., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **299**:77-82; Savage, T.J., Ichii, H., Hume, S.D., Little, D.B., and Croteau, R. (1994) *Arch. Biochem. Biophys.* **320**:257-265). Sequence alignment of 21 terpene synthases of plant origin (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024; Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268; Facchini, P.J., and Chappell, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**:11088-11092; Mau, C.J.D., and West, C.A. (1994) *Proc. Natl. Acad. Sci. USA* **91**:8497-8501; Yuba, A., Yazaki, K., Tabata, M., Honda, G., and Croteau, R. (1996) *Arch. Biochem. Biophys.* **332**:280-287; Wildung, M.R., and Croteau, R. (1996) *J. Biol. Chem.* **271**:9201-9204; Yamaguchi, S., Saito, T., Abe, H., Yamane, H., Murofushi, N., and Kamiya, Y. (1996) *Plant J.* **10**:203-213; Dudareva, N., Cseke, L., Blanc, V.M., and Pichersky, E. (1996) *Plant Cell* **8**:1137-1148; Chen, X.-Y., Chen, Y., Heinstein, P., and Davisson, V.J. (1995) *Arch. Biochem. Biophys.* **324**:255-266; Chen, X.-Y., Wang, M., Chen, Y., Davisson, J., and Heinstein, P. (1996) *J. Nat. Prod.* **59**:944-951; Back, K., and Chappell, J. (1995) *J. Biol. Chem.* **270**:7375-7381) reveals two absolutely conserved arginine residues, corresponding to Arg^{184} and Arg^{365} of pinene synthase (SEQ ID NO:4), one highly conserved cysteine residue (pinene synthase Cys^{543})(SEQ ID NO:4), and one highly conserved histidine residue (pinene synthase His^{186})(SEQ ID NO:4). The DDXXD (SEQ ID NO:45) sequence motif (pinene synthase Asp^{379}, Asp^{380} and Asp^{383}) (SEQ ID NO:4) is absolutely conserved in all relevant plant terpene synthases, as are several

other amino acid residues corresponding to Phe^{198}, Leu^{248}, Glu^{322}, Trp^{329}, Trp^{460} and Pro^{467} of pinene synthase (SEQ ID NO:4).

Amino acid sequences of the plant terpene synthases were compared with each other and with the deduced sequences of several sesquiterpene synthases cloned from microorganisms (Proctor, R.H., and Hohn, T.M. (1993) *J. Biol. Chem.* **268**:4543-4548; Back, K., and Chappell, J. (1995) *J. Biol. Chem.* **270**:7375-7381; Hohn, T.M., and Desjardins, A.E. (1992) *Mol. Plant-Microbe Interactions* **5**:249-256). As with all other plant terpene synthases, no significant conservation in primary sequence exists between the monoterpene synthases from Grand fir and the terpene synthases of microbial origin, except for the DDXXD (SEQ ID NO:45) sequence motif previously identified as a common element of all terpene synthases, and prenyltransferases which employ a related electrophilic reaction mechanism (Croteau, R., Wheeler, C.J., Cane, D.E., Ebert, R., and Ha, H.-J. (1987) *Biochemistry* **26**:5383-5389; Chen, A., Kroon, P.A., and Poulter, D.C. (1994) *Protein Sci.* **3**:600-607; McCaskill, D., and Croteau, R. (1997) *Adv. Biochem. Engineering Biotech.* **55**:108-146). The evidence is presently insufficient to determine whether extant plant and microbial terpene synthases represent divergent evolution from a common ancestor, which may also have given rise to the prenyltransferases, or whether these similar catalysts evolved convergently.

EXAMPLE 10

A Strategy For Cloning Certain Gymnosperm Monoterpene Synthases

The present invention includes myrcene synthase, (-)-limonene synthase and (-)-pinene synthase proteins, and nucleic acid molecules that encode myrcene synthase, (-)-limonene synthase and (-)-pinene synthase proteins. The amino acid sequence of each of the myrcene synthase, (-)-limonene synthase and (-)-pinene synthase proteins of the present invention each includes at least one of the amino acid sequence elements disclosed in Table 1.

Table 1

Amino Acids	Sequence	Orientation	Comments
1. 70-77	H S N (L, I, V) W D D D (SEQ ID NO: 46)	F only	HS makes a poor reverse primer
2. 148-153	A L D Y V Y (SEQ ID NO: 47)	F and R	
3. 306-312	E L A K L E F (SEQ ID NO: 48)	F and R	

4. 328-333	R W W K E S (SEQ ID NO: 49)	F and R	F primer uses 1st nt of Ser codon; R uses 1st two nts of Arg codon (rare only)
5. 377-383	(V, I, L) L D D M Y D (SEQ ID NO: 50)	F and R	
6. 377-383	(V, I, L) L D D L Y D (SEQ ID NO: 51)	F and R	Degeneracy of V/I/L at 377 too high for single primer
7. 377-383	(V, I, L) L D D I Y D (SEQ ID NO: 52)	F and R	
8. 543-549	C Y M K D (N, H) P (SEQ ID NO: 53)	R	F primer can also be constructed with this peptide but is too close to the 3' end to be useful

The numbers set forth in Table 1 for the first and last amino acid residue of each of the peptide sequences is the number of the corresponding amino acid residue in the amino acid sequence of the (-)-pinene synthase (SEQ ID NO:4) isolated from *Abies grandis*. Where a sequence of amino acid residues appears in brackets, *e.g.*, (L,I,V) in Table 1, the first amino acid residue within the brackets is the residue that appears in the (-)-pinene synthase amino acid sequence set forth in SEQ ID NO:4. The subsequent amino acid residues within the brackets represent other amino acid residues that commonly occur at the corresponding position in the amino acid sequence of other *Abies grandis* enzymes involved in terpene synthesis.

In Table 1, the letter "F" refers to the forward PCR reaction, *i.e.*, the PCR reaction which synthesizes the sense nucleic acid strand that encodes a gymnosperm monoterpene synthase. The letter "R" refers to the reverse PCR reaction, *i.e.*, the PCR reaction that synthesizes the antisense nucleic acid molecule that is complementary to the sense nucleic acid strand synthesized in the forward PCR reaction.

In order to clone nucleic acid molecules encoding myrcene synthase, (-)-limonene synthase and (-)-pinene synthase of the present invention, one or more oligonucleotide molecules corresponding to at least a portion of one of the amino acid sequences set forth in Table 1 can be used as a probe or probes with which to screen a genomic or cDNA library derived from one or more gymnosperm species. In this

context, the term "corresponding," or "correspond" or "corresponds," means that the oligonucleotide base sequence either a) encodes all or part of at least one of the amino acid sequences set forth in Table 1, or b) is complementary to a base sequence that encodes all or part of at least one of the amino acid sequences set forth in Table 1. The oligonucleotide probe(s) may contain a synthetic base, such as inosine, which can be substituted for one or more of the four, naturally-occurring bases, *i.e.*, adenine ("A"), guanine ("G"), cytosine ("C") and thymine ("T"). Thus, for example, the following oligonucleotide sequences "correspond" to the tripeptide sequence M M M: 5'ATGATGATG3' (sense orientation) (SEQ ID NO:54); 3'TACTACTAC5' (antisense orientation) (SEQ ID NO:55) and 3'IACIACIAC5' (SEQ ID NO:56).

One or more oligonucleotide sequence(s), corresponding to at least a portion of at least one of the amino acid sequences set forth in Table 1, can be used to screen a nucleic acid library in order to identify myrcene synthase, (-)-limonene synthase and (-)-pinene synthase clones of the present invention, according to methods well known to one of ordinary skill in the art. *See, e.g., Sambrook et al, supra.* The stringency of the hybridization and wash conditions during library screening in accordance with the present invention, utilizing one or more oligonucleotide sequence(s) corresponding to at least a portion of at least one of the amino acid sequences set forth in Table 1, is at least: for the hybridization step, 6X SSPE, 40-45°C, for 36 hours; for the wash step, 3X SSPE, 45°C, 3 X 15 minute washes. The presently preferred hybridization and wash conditions during library screening, utilizing one or more oligonucleotide sequence(s) corresponding to at least a portion of at least one of the amino acid sequences set forth in Table 1, in accordance with the present invention are: for the hybridization step, 6X SSPE, 40-45°C, for 36 hours; for the wash step, 0.1X SSPE, 65°C-70°C, 3 X 15 minute washes.

Examples of oligonucleotide sequences, corresponding to at least one of the amino acid sequences set forth in Table 1, that hybridize, under the foregoing hybridization and wash conditions, to the sense strands of the nucleic acid sequences of the present invention that encode myrcene synthase, (-)-limonene synthase or (-)-pinene synthase proteins are set forth in Table 2.

Table 2

Nucleic Acid Sequence	Corresponds to:
GTG TCG TTG GAG ACC CTG CTG CTG (SEQ ID NO:57)	SEQ ID No. 46

CGG GAG CTG ATG CAG ATG (SEQ ID NO:58)	SEQ ID No. 47
CTC GAG CGG TTC GAG CTC AAG (SEQ ID NO:59)	SEQ ID No. 48
GCC ACC ACC TTC CTC TCG (SEQ ID NO:60)	SEQ ID No. 49
GAG GAG CTG CTG TAC ATG CTG (SEQ ID NO:61)	SEQ ID No. 50
GAG GAG CTG CTG GAG ATG CTG (SEQ ID NO:62)	SEQ ID No. 51

Similarly, each of the myrcene synthase, (-)-limonene synthase and (-)-pinene synthase clones set forth in SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:5, or a portion thereof, may be used as a probe to screen a nucleic acid library in order to isolate monoterpene synthase clones of the present invention, according to methods well known to one of ordinary skill in the art. *See, e.g., Sambrook et al, supra.* The stringency of the hybridization and wash conditions during library screening in accordance with the present invention is at least: for the hybridization step, 6X SSPE buffer at 45°C to 50°C for 36 hours; for the wash step, 3X SSPE buffer at 50°C (three, fifteen minute washes). In accordance with the present invention, the presently preferred hybridization and wash conditions during library screening utilizing any of the gymnosperm monoterpene synthase clones set forth in SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:5, or a portion thereof, as probe are: for the hybridization step, 6X SSPE, 40-45°C, for 36 hours; for the wash step, 0.1X SSPE, 70°C-75°C, 3 X 15 minute washes.

Additionally, at least two oligonucleotide sequence(s), each corresponding to at least a portion of at least one of the amino acid sequences set forth in Table 1, can be used in a PCR reaction to generate a portion of a myrcene synthase, (-)-limonene synthase or (-)-pinene synthase clone of the present invention, which can be used as a probe to isolate a full-length clone of a myrcene synthase, (-)-limonene synthase or (-)-pinene synthase clone of the present invention. Thus, oligonucleotides that are useful as probes in the forward PCR reaction correspond to at least a portion of at least one of the amino acid sequences disclosed in Table 1 as having the "F" orientation. Conversely, oligonucleotides that are useful as probes in the reverse PCR reaction correspond to at least a portion of at least one of the amino acid sequences disclosed in Table 1 as having the "R" orientation. PCR reactions can be carried out

according to art-recognized PCR reaction conditions, such as the PCR reaction conditions set forth in Example 1 herein and as set forth in "PCR Strategies", M.A. Innis, D.H. Gelfand and J.J. Sninsky, eds., 1995, Academic Press, San Diego, CA (Chapter 14); "PCR Protocols: A Guide to Methods and Applications", M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White, eds., Academic Press, NY (1990). The presently preferred PCR reaction conditions are:

dNTPs	200 µM each
$MgCl_2$	5-7 mM
F and R primers	100 nM - 1µM each
Taq polymerase	1-2 units/reaction
cDNA template	10-100 ng/reaction

Buffers, PCR grade water, and Chill-out wax or mineral oil

The presently preferred thermocycler conditions are:

Denaturation	94^0 x 2 min	1 cycle
Denaturation	94^0 x 45 s	35 cycles
Annealing	42^0 - 55^0 x 45 s - 1 min	"
Polymerization	72^0 x 1-2 min	"
Adenylation	72^0 x 10 min	1 cycle

EXAMPLE 11

Cloning and Characterization of cDNA Clones Encoding (-)-Camphene Synthase, (-)-β-Phellandrene Synthase, Terpinolene Synthase and (-)-limonene/(-)-α-pinene synthase

Comparison of resin analysis and the products generated by in vitro assay of the corresponding native enzymes (Lewinsohn, E., Savage, T.J., Gijzen, M., and Croteau, R. (1993) *Phytochem. Anal.* **4**, 220-225; Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**, 267-273.), with the products of the available recombinant monoterpene synthases (Bohlmann, J., Steele, C.L., and Croteau, R. (1997) *J. Biol. Chem.* **272**, 21784-21792), reveals that myrcene synthase, limonene synthase and pinene synthase do not account for all of the monoterpene synthases of grand fir. Molecular cloning of the complete family of grand fir monoterpene synthases is essential for evaluation of the regulatory role of each gene

in constitutive and induced resin formation, can lead to the identification of useful genetic markers for resistance, provide the tools for engineering improved defense, and yield a more diverse set of recombinant catalysts for comparative mechanistic and structural study. Consequently, cDNA molecules encoding additional monoterpene synthases were isolated from a Grand fir cDNA library as described in this Example.

Additional Grand fir monoterpene synthases were cloned and analyzed as follows.

Substrates, Reagents and cDNA library. [1-^{3}H]Geranyl diphosphate (250 Ci/mol) (Croteau, R., Alonso, W.R., Koepp, A E., and Johnson, M.A. (1994) *Arch. Biochem. Biophys.* **309**, 184-192), [1-^{3}H]farnesyl diphosphate (125 Ci/mol) (Dehal, S.S., and Croteau, R. (1988) *Arch. Biochern. Biophys.* **261**, 346-356) and [1-^{3}H]geranylgeranyl diphosphate (120 Ci/mol) (LaFever, R.E., Stofer Vogel, B., and Croteau, R. (1994) *Arch. Biochem. Biophys.* **313**, 139-149) were prepared as described in the foregoing publications. Terpenoid standards were from the inventors' own collection. All other biochemicals and reagents were purchased from Sigma Chemical Co. or Aldrich Chemical Co., unless otherwise noted. Construction of the λZAP II cDNA library, using mRNA isolated from wounded grand fir sapling stems (Lewinsohn, E., Steele, C.L., and Croteau, R. (1994) *Plant Mol. Biol. Rep.* **12**, 20-25), was described previously in (Stofer Vogel, B., Wildung, M., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**, 23262-23268).

PCR-based probe generation. PCR was performed in a total buffer volume of 50 µl containing 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 5 mM $MgCl_2$, 200 µM of each dNTP, 2.5 units of Taq polymerase (BRL) and 5 µl of purified grand fir stem cDNA library phage as template (1.5 x 10^9 pfu/ml). Three different PCR mixtures were evaluated containing either 1-5 µM of each primer, 1-5 µM of primer E (SEQ ID NO:21) only, or 1-5 µM of primer G (SEQ ID NO:23) only. After a denaturing step at 94°C for 2 min, 35 cycles of amplification were performed employing the following temperature program using a Gradient 96 Robocycler (Stratagene): one min at 94°C, one min at each 2°C increment from 44 to 66°C, and two min at 72°C. The amplicons were analyzed by agarose gel electrophoresis (Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., pp. 7.20-7.43, Cold Spring Harbor, NY), and the products were extracted from the gel, ligated into pT7Blue (Novagen) and transformed into E. coli XL1-Blue cells. Plasmid DNA was prepared from individual transformants and the inserts were fully sequenced (DyeDeoxy Terminator Cycle Sequencing, Applied Biosystems). In addition to

previously described DNA fragments (Bohlmann, J., Steele, C.L., and Croteau, R. (1997) *J. Biol. Chem.* **272**, 21784-21792), a new insert sequence was identified and was designated as probe 7 (SEQ ID NO:63).

Library screening. For library screening, 200 ng of probe 7 (SEQ ID NO:63) was amplified by PCR, and the resulting amplicon was gel purified, randomly labeled with [α-^{32}P]dATP (Feinberg, A.P., and Vogelstein, B. (1984) *Anal. Biochem.* **137**, 216-267) and used to screen replica filters of 5 x 10^4 plaques from the wound-induced grand fir stem cDNA library plated on E. coli LE392. Hybridization was performed for 20 h at 55°C in 3 x SSPE and 0.1% SDS. Filters were washed three times for 10 min at 55°C in 3 x SSPE with 0.1% SDS and exposed for 17 h to Kodak XAR film at –70°C (Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., pp. 7.20-7.43, Cold Spring Harbor, NY). Forty λZAPII clones yielding positive signals with probe 7 (SEQ ID NO:63) were purified through a second round of hybridization. Purified λZAPII clones were excised in vivo as Bluescript II SK$^-$ phagemids and transformed into E. coli XLOLR according to the manufacturer's instructions (Stratagene). The size of each cDNA insert was determined by PCR using T3 (SEQ ID NO:29) and T7 (SEQ ID NO:30) promoter primers, and size-selected inserts (>1.5 kb) were partially sequenced from both ends to reveal several acquisitions of four unique clones, including apparent full-length versions of three (designated 7.30 (ag6) (SEQ ID NO:64), 7.36 (ag8) (SEQ ID NO:66) and 7.32 (ag11) (SEQ ID NO:68)), and a fourth (7.31) (SEQ ID NO:70) that was clearly 5'-truncated.

Rapid amplification of cDNA ends. To acquire the 5'-terminus corresponding to truncated cDNA clone ag7.31 (SEQ ID NO:70), 5'-rapid amplification of cDNA ends (RACE) was carried out using the Marathon cDNA amplification system (Clontech) and the manufacturer's protocol. Nested reverse RACE primers specific for ag7.31 (SEQ ID NO:70), designated 10-2 (5'-ACG AAG CTT CTT CTC CAC GG-3')(SEQ ID NO:71) and 10-4 (5'-GGA TCC CAT CTC TTA ACT GC-3')(SEQ ID NO:72), were used in combination with primer AP1 (SEQ ID NO:73) and AP2 (SEQ ID NO:74) (Clontech), respectively. The resulting amplicon was cloned into vector pT7Blue (Novagen) and completely sequenced on both strands. The full length cDNA corresponding to clone ag7.31 (SEQ ID NO:70) was then amplified by PCR using primers AG9F (5'-ATG GCT CTT GTT TCT ATC TTG CCC-3') (SEQ ID NO:75) and AG9R (5'-TTA CAA AGG CAC AGA CTC AAG GAC-3')(SEQ ID

NO:76), ligated into pT7Blue and the resulting plasmid was designated pAG9 (SEQ ID NO:77).

The foregoing similarity based cloning approach was designed to employ PCR primers E (SEQ ID NO:21) and G (SEQ ID NO:23) to generate new terpenoid synthase cDNA fragments for isolation and deployment as hybridization probes in screening an enriched library. By using wound-induced grand fir stem library cDNA as template, the combination of primers E (SEQ ID NO:21) and G (SEQ ID NO:23) amplified primarily a 1022 bp PCR product identical to probe 3 (SEQ ID NO:24) and an additional, very minor product of 296 bp. This very low abundance 296 bp product was also amplified when only primer E (SEQ ID NO:21) was used in the PCR reaction with the Grand fir cDNA library template in the absence of a second primer. Cloning and sequencing of this PCR product (designated probe 7) (SEQ ID NO:63) unexpectedly revealed significant deduced sequence similarity with previously cloned grand fir terpene synthases in the region corresponding to amino acid Val^{424}-Leu^{521} of (-)-pinene synthase (SEQ ID NO:4). The highest levels of identity were observed at the nucleic acid level with monoterpene synthase clones ag2.2 (SEQ ID NO:1) (myrcene synthase, 74%), ag3.18 (SEQ ID NO:3) (-)-pinene synthase, 69%) and ag10 (SEQ ID NO:5) (-)-limonene synthase, 83%). Lower levels of similarity were noted with the sesquiterpene synthases ag1 (50%), ag4 (54%) and ag5 (57%), and still less with the diterpene synthase abietadiene synthase (46%), suggesting that this PCR product represented a fragment of a new monoterpene synthase. In retrospect, generation of this fragment can be explained based on two unexpected annealing events of primer E (SEQ ID NO:21) to a monoterpene synthase cDNA in the regions corresponding to amino acids Val^{424}-Ala^{432} and Leu^{513}-Leu^{521} of (-)-pinene synthase (SEQ ID NO:4). These binding sites for primer E (SEQ ID NO:21) are significantly different from the amino acid sequence GE(K/T)(V/I)M(E/D)EA (SEQ ID NO:26) to which the degenerate primer was designed. With nucleic acid sequence identities between primer E (SEQ ID NO:21) (23 nucleotides) and the corresponding two sites of recognition in previously acquired grand fir monoterpene synthases in the range of 44-57% and 48-56%, respectively, amplification of probe 7 (SEQ ID NO:63) would not be predicted. Probe 7 (SEQ ID NO:63) was subsequently employed for filter hybridization of the wound induced grand fir phage cDNA library.

Screening of 10^5 cDNA phage plaques from the wounded grand fir stem library with probe 7 (SEQ ID NO:63) yielded 40 positives of which 37 were isolated by one additional round of filter hybridization, excised in vivo and partially sequenced

from both ends. Sequence analysis revealed four new, unique cDNA fragments represented by phage clones 7.30 (SEQ ID NO:64), 7.36 (SEQ ID NO:66) 7.31 (SEQ ID NO:70) and 7.32 (SEQ ID NO:68), of which clone 7.32 (SEQ ID NO:68) was identical to probe 7 (SEQ ID NO:63). Complete sequencing of these inserts and sequence comparison placed these genes into the gymnosperm Tpsd subfamily of plant terpenoid synthases (Bohlmann, J., Meyer-Gauen, G., and Croteau. R. (1998) *Proc. Natl. Acad. Sci.* USA **95**, 4126-4133) with closest relationship to previously acquired grand fir monoterpene synthases. Alignment of the deduced amino acid sequences of the four new, presumptive terpene synthase fragments with extant monoterpene synthases indicated that clones 7.30 (SEQ ID NO:64), 7.36 (SEQ ID NO:66) and 7.32 (SEQ ID NO:68) represented full-length versions, whereas clone 7.31 (SEQ ID NO:70) was truncated at the 5'-terminus. A full-length clone corresponding to truncated cDNA 7.31 (SEQ ID NO:70) was obtained by a 5'-RACE method. These full-length clones were designated ag6 (7.30) (SEQ ID NO:64), ag8 (7.36) (SEQ ID NO:66), ag9 (7.31) (SEQ ID NO:77) and ag11 (7.32) (SEQ ID NO:68), consistent with the nomenclature for other terpenoid synthases from grand fir (Bohlmann, J., Meyer-Gauen, G., and Croteau. R. (1998) *Proc. Natl. Acad. Sci.* USA 95, 4126-4133).

Sequence analysis. Inserts of all recombinant Bluescript plasmids and pSBET plasmids were completely sequenced on both strands via primer walking using the DyeDeoxy Terminator Cycle Sequencing method (Applied Biosystems). Sequence analysis was conducted using programs from the Genetics Computer Group (Genetics Computer Group (1996) Program Manual for the Wisconsin Package, Version 9.0, Genetics Computer Group, Madison, WI).

The sequences of clone ag6 (2013 bp with ORF of 1854 nt encoding 618 amino acids) (SEQ ID NO:64), clone ag8(2186 bp with ORF of 1890 nt encoding 630 amino acids) (SEQ ID NO:66), clone ag9 (1893 bp with ORF of 1890 nt encoding 630 amino acids) (SEQ ID NO:77) and clone ag11 (2429 bp with ORF of 1911 nt encoding 637 amino acids) (SEQ ID NO:68) revealed, in addition to overall similarities, other features characteristic of monoterpene synthases. The lengths of the deduced proteins (618 – 637 amino acids) and the predicted molecular weights (71,000 – 73,000) are in the range of other monoterpene synthases. The deduced proteins are larger than the sesquiterpene synthases, δ-selinene synthase (581 amino acids) and γ-humulene synthases (593 amino acids) (Steele, C.L., Crock, J., Bohlmann, J., and Croteau, R. (1998) *J. Biol. Chem.* 273, 2078-2089) but smaller

than abietadiene synthase (868 amino acids) (Stofer Vogel, B., Wildung, M., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**, 23262-23268) and (E)-α-bisabolene synthase (817 amino acids) (Bohlmann, J., Crock, J., Jetter, R., and Croteau, R. (1998) *Proc. Natl. Acad. Sci.* USA **95**, 6756-6761) from grand fir. As with other monoterpene synthases (Bohlmann, J., Meyer-Gauen, G., and Croteau. R. (1998) *Proc. Natl. Acad. Sci.* USA **95**, 4126-4133), ag6 (SEQ ID NO:64), ag8 (SEQ ID NO:66), ag9 (SEQ ID NO:77) and ag11 (SEQ ID NO:68) appear to encode preproteins bearing an amino-terminal transit peptide for plastidial import of these nuclear gene products (Williams, D.C., McGarvey, D.J., Katahira, E.J., and Croteau, R. (1998) *Biochemisiry* **37**, 12213-12220; Turner, G., Gershenzon, J., Nielson, E.E., Froehlich, J.E., and Croteau, R. (1999) *Plant Physiol.*, in press). In all cases, the amino-terminal 50-60 residues of the deduced sequences are rich in serine (15-21%) but have few acidic residues, consistent with such targeting peptides (Keegstra, K., Olsen, J.J., and Theg, S.M. (1989) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 471-501; von Heijne, G., Steppuhn, J., and Herrmann, R.G. (1989) *Eur. J. Biochem.* **180**, 535-545). A tandem arginine element has recently been demonstrated to approximate the putative amino-terminus of mature monoterpene synthases, and to be involved in the initial diphosphate migration step of the coupled monoterpene cyclization reaction sequence in which geranyl diphosphate is isomerized to enzyme-bound linalyl diphosphate (Williams, D.C., McGarvey, D.J., Katahira, E.J., and Croteau, R. (1998) *Biochemisiry* **37**, 12213-12220). The position of the tandem arginines is similar in ag6 (SEQ ID NO:65) ($R^{59}R^{60}$), ag8 (SEQ ID NO:67) ($R^{62}R^{63}$), ag9 (SEQ ID NO:78) ($R^{65}R^{66}$) and ag11 (SEQ ID NO:69) ($R^{71}R^{72}$), and is conserved in all monoterpene synthases of angiosperm and gymnosperm origin in a position nine amino acids upstream of an absolutely conserved tryptophan residue (Bohlmann, J., Meyer-Gauen, G., and Croteau. R. (1998) *Proc. Natl. Acad. Sci.* USA **95**, 4126-4133). All other previously described motifs of plant terpene synthases of the tpsa, tpsb and tpsd subfamilies (Bohlmann, J., Meyer-Gauen, G., and Croteau. R. (1998) *Proc. Natl. Acad. Sci.* USA **95**, 4126-4133), including the aspartate-rich DDXXD (SEQ ID NO:45) element involved in coordinating the divalent metal ion for substrate binding (Ashby, M.N., and Edwards, P.A. (1 990) *J. Biol. Chem.* **265**, 13157-13164; Marrero, P.F., Poulter, C.D., and Edwards, P.A. (1992) *J. Biol. Chem.* **267**, 21873-21878; Tarshis, L.C., Yan, M., Poulter, C.D., and Sacchettini, J.C. (1994) *Biochemistry* **33**, 10871-10877; Cane, D.E., Sohng, J.K., Lamberson, C.R., Rudnicki, S.M.. Wu, Z., Lloyd, M.D., Oliver, J.S., and Hubbard, B.R. (1994) *Biochemistry* **33**,

5846-5857), are also found in the enzymes encoded by ag6 (SEQ ID NO:65), ag8 (SEQ ID NO:70), ag9 (SEQ ID NO:78) and ag11 (SEQ ID NO:69).

cDNA expression in E. coli and enzyme assays. The fully sequenced insert fragments of plasmids pAG6 (SEQ ID NO:64), pAG8 (SEQ ID NO:66), pAG9 (SEQ ID NO:77) and pAG11 (SEQ ID NO:68) were subcloned in-frame into the expression vector pSBETa (Schenk, P.M., Baumann, S., Mattes, R., and Steinbiss, H. (1995) *BioTechniques* **19**, 196-200). First, internal NdeI sites and internal BamHI sites in cDNA clones pAG6 (SEQ ID NO:64), pAG9 (SEQ ID NO:77) and pAG11 (SEQ ID NO:68) were eliminated by site directed mutagenesis using the Quick Change Mutagenesis system (Stratagene). Mutagenesis primers designated 6eBamHIF (5'-CAA TTA AGA GAT GGG ACC CGT CCG CGA TGG-3') (SEQ ID NO:79) and 6eBamHIR (5'-CCA TCG CGG ACG GGT CCC ATC TCT TAA TTG-3') (SEQ ID NO:80) were used to eliminate an internal BamHI site in pAG6 (SEQ ID NO:64), 9eBamHIF (5'-GCA TTT AAG AGA TGG GAC CCG TCT GCC ACA G-3') (SEQ ID NO:81) and 9eBamHIR (5'-CTG TGG CAG ACG GGT CCC ATC TCT TAA ATG C-3') (SEQ ID NO:82) to eliminate an internal BamHI site in pAG9 (SEQ ID NO:77), and 732eNdeIF (5'-CGA GAT GCC ATA CGT GAA TAC GCA G-3') (SEQ ID NO:83) and 732eNde1R (5'-CTG CGT ATT CAC GTA TGG CAT CTC G-3') (SEQ ID NO:84) to eliminate an internal NdeI site in pAG11 (SEQ ID NO:68).

To introduce suitable restriction sites for subcloning full-length versions and versions truncated to remove the presumptive plastidial targeting peptide of these enzymes (Bohlmann, J., Meyer-Gauen, G., and Croteau. R. (1998) *Proc. Natl. Acad. Sci.* USA **95**, 4126-4133; Williams, D.C., McGarvey, D.J., Katahira, E.J., and Croteau, R. (1998) *Biochemisiry* **37**, 12213-12220), fragments were amplified by PCR using primer 6-NdeI-M (5'-CTG ATA GCA AGC TCA TAT GGC TCT TCT TTC-3') (SEQ ID NO:85) or primer 6-NdeI-R (5'-GCC CAC GCG TCT CAT ATG AGA ATC AGT AGA TGC G-3') (SEQ ID NO:86) individually in combination with primer 6-BamHI (5'-CAC CCA TAG GGG ATC CTC AGT TAA TAT TG-3') (SEQ ID NO:87) for pAG6 (SEQ ID NO:64), primer 8-NdeI-M (5'-TAA GCG AGC ACA TAT GGC TCT GGT TTC TTC-3') (SEQ ID NO:88) in combination with primer 8-BamHI (5'-GCA TAA ACG CAT AGC GGA TCC TAC ACC AA-3') (SEQ ID NO:89) for pAG8 (SEQ ID NO:66), primer 9-NdeI-M (5'-CCC GGG GAT CGG ACA TAT GGC TCT TGT TTC-3') (SEQ ID NO:90) in combination with primer 9-BamHI (5'-GGT CGA CTC TAG AGG ATC CAC TAG TGA TAT GGA T-3') (SEQ ID NO:91) for pAG9 (SEQ ID NO:77), and primer 11-NdeI-M (5'-GAA CAT

ATG GCT CTC CTT TCT ATC GTA-3') (SEQ ID NO:92) or primer 11-NdeI-R (5'-GGT GGT GGT GTA CAT ATG AGA CGC ATA CGG G-3') (SEQ ID NO:93) in combination with primer 11-BamHI (5'-GAG ACT AGA CTG GAT CCC ATA TAC ACT GTA ATG G-3') (SEQ ID NO:94) for pAG11 (SEQ ID NO:68). PCR reactions were performed in volumes of 50 µl containing 20 mM Tris-HCl (pH 8.8), 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100, 5 µg bovine serum albumin, 200 µM of each dNTP, 0.1 µM of each primer, 2.5 units of recombinant Pfu DNA polymerase (Stratagene) and 100 ng plasmid DNA using the following program: denaturation at 94°C for 1 min, annealing at 60°C for 1 min, extension at 72°C for 3.5 min; 35 cycles with final extension at 72°C for 5 min. The resulting PCR products were purified by agarose gel electrophoresis and employed as templates for a second PCR amplification under the identical conditions but in a total volume of 250 µl each. Products from this secondary amplification were digested with the above indicated restriction enzymes, purified by ultrafiltration, and then ligated into NdeI/BamHI-digested pSBETa to yield the respective plasmids pSAG6(M), pSAG6(R), pSAG8(M), pSAG9(M), pSAG11(M) and pSAG11(R).

For use as controls in the expression of full-length and truncated forms, inserts of the previously described (Bohlmann, J., Steele, C.L., and Croteau, R. (1997) *J. Biol. Chem.* **272**, 21784-21792) plasmids pGAG2.2 (myrcene synthase) (SEQ ID NO:1) and pGAG3.18 ((-)-pinene synthase) (SEQ ID NO:3) were prepared for subcloning into pSBETa by PCR amplification using primer 2-NdeI-M (5'-CAA AGG GAG CAC ATA TGG CTC TGG-3') (SEQ ID NO:95) or primer 2-NdeI-R (5'-CTG ATG ATG GTC ATA TGA GAC GCA TAG GTG-3') (SEQ ID NO:96) in combination with primer 2-BamHI (5'-GAC CTT ATT ATT ATG GAT CCG GTT ATA G-3') (SEQ ID NO:97) for pGAG2.2 (SEQ ID NO:1), and primer 3-NdeI-R (5'-CCG ATG ATG GTC ATA TGA GAC GCA TGG GCG-3') (SEQ ID NO:98) in combination with primer 3-BamHI (5'-GGG CAT AGA TTT GAG CGG ATC CTA CAA AGG-3') (SEQ ID NO:99) for pGAG3.18 (SEQ ID NO:3). Prior to insert subcloning, two internal BamHI sites and two internal NdeI sites were eliminated from the insert of pGAG3.18 (SEQ ID NO3:) by site directed mutagenesis using primers 3e1BamHIF (5'-CGT TTG GGA ATC CAT AGA CAT TTC-3') (SEQ ID NO:100), 3elBamHIR (5'-GAA ATG TCT ATG GAT TCC CAA ACG-3') (SEQ ID NO:101), 3e2BamHIF (5'-GAA GAG ATG GGA CCC GTC CTC GAT AG-3') (SEQ ID NO:102), 3e2BamHIR (5'-CTA TCG AGG ACG GGT CCC ATC TCT TC-3') (SEQ ID NO:103), 3elNdeIF (5'-GAA CAC GAA GTC CTA TGT GAA

GAG C-3') (SEQ ID NO:104), 3elNdeIR (5'-GCT CTT CAC ATA GGA CTT CGT GTT C-3') (SEQ ID NO:105), 3e3NdeIF (5'-GAT ACG CTC ACT TAT GCT CGG GAA G-3') (SEQ ID NO:106) and 3e2NdeIR (5'-CTT CCC GAG CAT AAG TGA GCG TAT C-3') (SEQ ID NO:107). Subcloning into pSBETa yielded pSAG2(M) and pSAG2(R) for myrcene synthase, and pSAG3(R) for (-)-pinene synthase. All recombinant pSBETa plasmids were confirmed by sequencing to insure that no errors had been introduced by the polymerase reactions, and were then transformed into E. coli BL21(DE3) by standard methods.

For functional expression, bacterial strains E. coli BL21(DE3)/pSAG2(M), E. coli BL21(DE3)/pSAG2(R), E. coli BL21(DE3)/pSAG3(R), E. coli BL21(DE3)/pSAG6(M), E. coli BL21(DE3/pSAG6(R), E. coli BL21(DE3)/pSAG8(M), E. coli BL21(DE3)/pSAG9(M), E. coli BL21(DE3)/pSAG11(M) and E. coli BL21(DE3)/SAG11(R) were grown to A_{600} = 0.5 at 37°C in 5 ml of Luria-Bertani medium (Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., pp. 7.20-7.43, Cold Spring Harbor, NY) supplemented with 30 µg kanamycin/ml. Cultures were then induced by addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside and grown for another 12 h at 20°C. Cells were harvested by centrifugation and disrupted by sonication followed by centrifugation, and the resulting soluble enzyme preparations were assayed for monoterpene, sesquiterpene and diterpene synthase activity using the appropriate radiolabeled substrate as described previously (Bohlmann, J., Steele, C.L., and Croteau, R. (1997) *J. Biol. Chem.* **272**, 21784-21792). This assay involves the isolation and separation of terpene olefins and oxygenated terpenes by column chromatography on silica, followed by scintillation counting to determine conversion rate. GC-MS analysis is employed for product identification, by comparison of retention times and mass spectra to those of authentic standards (Bohlmann, J., Steele, C.L., and Croteau, R. (1997) *J. Biol. Chem.* **272**, 21784-21792); assignment of absolute configuration is based upon chiral column capillary GC and matching of retention time to that of the corresponding enantiomer (Lewinsohn, E., Savage, T.J., Gijzen, M., and Croteau, R. (1993) *Phytochem. Anal.* **4**, 220-225). To insure production of sufficient material for identification, the standard assay (Bohlmann, J., Steele, C.L., and Croteau, R. (1997) *J. Biol. Chem.* **272**, 21784-21792) was scaled up by a factor of 20.

For functional characterization, cDNAs ag6 (SEQ ID NO:64), ag8 (SEQ ID NO:66), ag9 (SEQ ID NO:77) and ag11 (SEQ ID NO:68) were subcloned into the

bacterial expression vector pSBET (Schenk, P.M., Baumann, S., Mattes, R., and Steinbiss, H. (1995) *BioTechniques* **19**, 196-200). This vector employs the T7 RNA polymerase promoter and contains the argU gene for expression in E. coli of the tRNA using the rare arginine codons AGA and AGG that are commonly found in plant genes. pSBET constructs have been employed previously for successful bacterial expression of terpene synthases from gymnosperms and angiosperms (Steele, C.L., Crock, J., Bohlmann, J., and Croteau, R. (1998) *J. Biol. Chem.* 273, 2078-2089; Bohlmann, J., Steele, C.L., and Croteau, R. (1997) *J. Biol. Chem.* **272**, 21784-21792; Williams, D.C., McGarvey, D.J., Katahira, E.J., and Croteau, R. (1998) *Biochemisiry* **37**, 12213-12220).

Initially, the four pSBET plasmids pSAG6(M), pSAG8(M), pSAG9(M) and pSAG11(M), containing the full-length cDNA inserts of ag6 (SEQ ID NO:64), ag8 (SEQ ID NO:66), ag9 (SEQ ID NO:77) and ag11 (SEQ ID NO:68), respectively, were expressed in E. coli BL21(DE3). Extracts of the induced, transformed bacterial cells were assayed for monoterpene, sesquiterpene and diterpene synthase activity using the corresponding labeled C_{10}, C_{15} and C_{20} prenyl diphosphate substrates (Bohlmann, J., Steele, C.L., and Croteau, R. (1997) *J. Biol. Chem.* **272**, 21784-21792). Enzymatic production of terpene olefin(s) was observed only with E. coli strains BL21(DE3)/pSAG8(M) and BL21(DE3)/pSAG9(M) using [1-^{3}H]geranyl diphosphate as substrate. Since it is known that the relatively large amino-terminal transit peptide of plant terpene synthases can impair expression of the functional preprotein in E. coli and promote the formation of intractable inclusion bodies (Williams, D.C., McGarvey, D.J., Katahira, E.J., and Croteau, R. (1998) *Biochemisiry* **37**, 12213-12220), the "pseudomature" forms of ag6 (SEQ ID NO:64) and ag11 (SEQ ID NO:68) were prepared by truncation of the cDNAs to insert a starting methionine immediately upstream of their tandem arginine elements ($R^{59}R^{60}$ for ag6 (SEQ ID NO:64), $R^{71}R^{72}$ for ag11 (SEQ ID NO:68)) for expression in E. coli BL21(DE3)/pSAG6(R) and E. coli BL21(DE3)/pSAG11(R). Similar truncations of previously characterized ag2.2 (SEQ ID NO:1) (myrcene synthase, $R^{64}R^{65}$) and ag3.18 (SEQ ID NO:3) ((-)-pinene synthase, (SEQ ID NO:5) $R^{64}R^{65}$) (Bohlmann, J., Steele, C.L., and Croteau, R. (1997) *J. Biol. Chem.* **272**, 21784-21792) were prepared for expression from pSBET as controls for the possible alteration of product yield and distribution resulting from truncation. These three latter constructs, pSAG2(M), pSAG2(R) and pSAG3(R), afforded high level expression of activity in BL21(DE3) cells, and revealed the identical product patterns previously observed for

the myrcene synthase and pinene synthase preprotein forms expressed from pBluescript and pGEX vectors (Bohlmann, J., Steele, C.L., and Croteau, R. (1997) *J. Biol. Chem.* **272**, 21784-21792). Thus, truncation directly upstream of the arginine pair does not alter product formation of these recombinant monoterpene synthases. When ag6 (SEQ ID NO:64) and ag11 (SEQ ID NO:68) were expressed as their truncated forms pSAG6(R) and pSAG11(R) in E. coli BL21(DE3), and the extracts assayed for all three terpene synthase activities, only geranyl diphosphate yielded high levels of terpene olefin products.

Monoterpenes generated at preparative scale from geranyl diphosphate by the recombinant enzymes encoded by ag6 (SEQ ID NO:64), ag8 (SEQ ID NO:66), ag9 (SEQ ID NO:77) and ag11 (SEQ ID NO:68) were analyzed by GC-MS and chiral phase (β-cyclodextrin) capillary GC, and identified by comparison of retention times and mass spectra to authentic standards (FIGURES 6-9). The synthase encoded by ag6 (SEQ ID NO:65) was shown to produce three principal products (FIGURE 6A), the major one of which was identified as (-)-1S,4R-camphene (54%) (FIGURES 6B and C), followed by (-)-1S,5S-α-pinene (32%) and (-)-4S-limonene (7%) which were identified by similar means (data not shown (These additional data may be accessed at website www.wsu.edu/~ibc/faculty/rc.html)). The enzyme encoded by ag6 (SEQ ID NO:65) is therefore designated as (-)-1S,4R-camphene synthase. Interestingly, the (-)-camphene synthase (SEQ ID NO:65) is not as stereoselective as other monoterpene synthases in producing both (-)-α-pinene and (+)-α-pinene as coproducts in a ratio of 95:5; the (-)-enantiomer dominates to the extent of more than 99% in this pinene isomer formed by (-)-pinene synthase from grand fir (Bohlmann, J., Steele, C.L., and Croteau, R. (1997) *J. Biol. Chem.* **272**, 21784-21792). Soluble enzyme preparations from E. coli BL21(DE3)/pSAG8(M) converted geranyl diphosphate to four products (FIGURE 7A) with β-phellandrene (52%) as the major olefin (FIGURE 7B and C), and lesser amounts of (-)-1S,5S-β-pinene (34%), (-)-1S,5S-α-pinene (8.5%) and (-)-4S-limonene (6%) identified by similar means (data not shown (these additional data may be accessed at website www.wsu.edu/~ibc/faculty/rc.html)). The stereochemistry of β-phellandrene was not confirmed directly, since only the authentic (+)-enantiomer was available as a standard for chiral phase GC analysis. Nevertheless, stereochemical considerations based on the established absolute configuration of the co-products, and the natural occurrence of (-)-4S-β-phellandrene in the turpentine (Lewinsohn, E., Savage, T.J., Gijzen, M., and Croteau, R. (1993) *Phytochem. Anal.* **4**, 220-225), suggest that the biosynthetic

product is the mechanistically related (-)-4S-antipode derived via the (+)-3S-linalyl diphosphate intermediate (Croteau, R. (1 9 87) *Chem. Rev.* **87**, 929-954; Wise, M.L., and Croteau, R. (1998) *in* Comprehensive Natural Products Chemistry: Isoprenoids (Cane, D.E., Ed.), Vol. 2, pp. 97-153, Elsevier Science, Oxford). The product of the ag8 gene (SEQ ID NO:67) is therefore designated (-)-4S-β-phellandrene synthase.

The enzyme encoded by cDNA clone ag9 (SEQ ID NO:78) also produced several monoterpenes from geranyl diphosphate (FIGURE 8), with the achiral olefin terpinolene identified as the major product (42%) (FIGURE 8B and C). In addition to minor amounts of four unidentified olefins (FIGURE 8A), (-)-α-pinene (18%), (-)-limonene (11%) and (-)-β-pinene (10%) were also identified as significant biosynthetic products by mass spectrometric and chromatographic analysis (data not shown (These additional data may be accessed at website www.wsu.edu/~ibc/faculty/rc.html)), indicating that the entire product set was derived in stereochemically consistent fashion via (+)-3S-linalyl diphosphate as intermediate (Croteau, R. (1 9 87) *Chem. Rev.* **87**, 929-954; Wise, M.L., and Croteau, R. (1998) *in* Comprehensive Natural Products Chemistry: Isoprenoids (Cane, D.E., Ed.), Vol. 2, pp. 97-153, Elsevier Science, Oxford). The enzyme encoded by ag9 (SEQ ID NO:78) is designated terpinolene synthase based on the principal olefinic product. cDNA clone ag11 (SEQ ID NO:68) also encodes a multiple product monoterpene synthase (FIGURE 9) with the two most abundant products identified as (-)-4S-limonene (35%) and (-)-1S,5S-α-pinene (24%) (FIGURE 9B and C), with lesser amounts of β-phellandrene (20%, presumably also the (-)-antipode), (-)-1S,5S-β-pinene (11%), and (-)-1S,5S-sabinene (10%) (data not shown (These additional data may be accessed at website www.wsu.edu/~ibc/faculty/rc.html)). This is the first recombinant monoterpene synthase to produce such a range of cyclic olefins, none of which truly dominate the profile (FIGURE 9A). Based on the close sequence relatedness between clone ag11 (SEQ ID NO:68) and the previously described (-)-limonene synthase of grand fir (ag10 (SEQ ID NO:5), which produces about 80% (-)-limonene (Bohlmann, J., Steele, C.L., and Croteau, R. (1997) *J. Biol. Chem.* **272**, 21784-21792)), ag11 (SEQ ID NO:69) was previously considered to be a likely (-)-limonene synthase. However, given the significant production of (-)-α-pinene by this cyclase, unlike the ag10 (-)-limonene synthase (SEQ ID NO:6), it seems appropriate to designate the product of the ag11 gene (SEQ ID NO:69) as (-)-limonene/(-)-α-pinene synthase to clearly distinguish the two. The (-)-limonene/(-)-α-pinene synthase (ag11) (SEQ ID NO:69) and the (-)-limonene

synthase (ag10) (SEQ ID NO:6) share 93% similarity and 91% identity at the deduced amino acid level, demonstrating that less than 10% sequence divergence is sufficient to result in a significantly different product outcome. Interestingly, all sequence differences between ag11 (SEQ ID NO:69) and ag10 (SEQ ID NO:6) are confined to the carboxy-terminal half of these proteins, which is thought to comprise the active site region involved in the cyclization step(s) catalyzed by sesquiterpene synthases and presumably other terpene synthases as well (Starks, C.M., Back, K., Chappell, J., and Noel, J.P. (1997) *Science* **277**, 1815-1820). This finding is relevant to the rational redesign of terpene synthases, and to the evolution of these catalysts which almost certainly involves gene duplication and modification as the basis for generating new terpenoid synthase function (Bohlmann, J., Meyer-Gauen, G., and Croteau. R. (1998) *Proc. Natl. Acad. Sci.* USA **95**, 4126-4133).

cDNAs encoding (-)-camphene synthase, (-)-β-phellandrene synthase, terpinolene synthase and (-)-limonene/(-)-α-pinene synthase have not been described previously from any source. Together with the cloned myrcene synthase (SEQ ID NO:1), (-)-limonene synthase (SEQ ID NO:5) and (-)-pinene synthase (SEQ ID NO:3), these seven enzymes account for the production of most, but not all, monoterpenes of the constitutive turpentine (Lewinsohn, E., Savage, T.J., Gijzen, M., and Croteau, R. (1993) *Phytochem. Anal.* **4**, 220-225) and wound-induced resin (Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**, 267-273) of grand fir. The cDNA encoding a constitutive and/or inducible (+)-3-carene synthase (Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**, 267-273; Savage, T.J., and Croteau, R. (1993) *Arch. Biochem. Biophys.* **305**, 581-587) has thus far eluded detection. Based on analytical results (Lewinsohn, E., Savage, T.J., Gijzen, M., and Croteau, R. (1993) *Phytochem. Anal.* **4**, 220-225) and in vitro assays (Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**, 267-273), the β-phellandrene synthase (SEQ ID NO:67) contributes principally to production of the constitutive turpentine. The terpinolene synthase (SEQ ID NO:78) may also represent a primarily constitutive enzyme; however, the complex and overlapping product profiles of these synthases do not allow unambiguous assignment of their functional role(s) in constitutive resin synthesis or the induced response without more detailed RNA blot analysis. A recent analytical survey of a small grand fir population for different chemotypes in both constitutive resin production and the induced response indicates considerable variation in the constitutive and inducible deployment of the members of this tpsd

multiple gene family (Katoh, S., and Croteau, R. (1998) *Phytochemistry* **47**, 577-582).

EXAMPLE 12

Hybridization Conditions

Presently preferred nucleic acid molecules of the present invention hybridize under stringent conditions to either one or both of hybridization probe A and hybridization probe B (or to their complementary antisense sequences). Hybridization probe A has the nucleic acid sequence of the portion of SEQ ID NO:3 extending from nucleotide 1560 to nucleotide 1694. Hybridization probe B has the nucleic acid sequence of the portion of SEQ ID NO:5 extending from nucleotide 1180 to nucleotide 1302. High stringency conditions are defined as hybridization in 5x SSC at 65°C for 16 hours, followed by two washes in 2x SSC at 20°C to 26°C for fifteen minutes per wash, followed by two washes in 0.2x SSC at 65°C for twenty minutes per wash. Moderate stringency conditions are defined as hybridization in 3x SSC at 65°C for 16 hours, followed by two washes in 2x SSC at 20°C to 26°C for twenty minutes per wash, followed by one wash in 0.5x SSC at 55°C for thirty minutes. Low stringency conditions are defined as hybridization in 3x SSC at 65°C for 16 hours, followed by two washes in 2x SSC at 20°C to 26°C for twenty minutes per wash.

EXAMPLE 13

Characteristics of Presently Preferred Monoterpene Synthase Proteins of the Invention

Presently preferred monoterpene synthase proteins of the present invention have the following characteristics: a preprotein molecular weight of from about 71 kiloDaltons (kDa) to about 73 kDa; a mature protein (excluding plastidial transit peptide) molecular weight of from about 60 kDa to about 67 kDa; isoelectric point (pI) of from about 4.5 to about 5.0 as determined by isoelectric focussing; a mature protein pH optimum of from about 7.5 to about 8.0; utilization of geranyl diphosphate as a substrate with K_m less than 15μM; a requirement for a divalent metal ion as a cofactor, with manganese (Mn^{2+}) being preferred over magnesium (Mg^{2+}), the Km for binding of Mn^{2+} being less than 50μM; biological activity is enhanced by the presence of K^+ ions at a concentration of 100μM; the presence of the sequence motif DDXXD (SEQ ID NO:45) within the carboxy terminal half of the protein, and the presence of two tandem arginines immediately following the C-terminal end of the plastidial transit peptide. The foregoing characterisitics apply to the mature protein, unless otherwise stated.

EXAMPLE 14

Alteration of Monoterpene Levels and Composition in Plant Seeds

In accordance with the present invention, methods for increasing production of monoterpene compounds in a plant, particularly in plant seeds, are provided. The methods involve transforming a plant cell with a nucleic acid sequence encoding at least one monoterpene synthase, such as those encoded by the nucleic acid sequences set forth in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO64:, SEQ ID NO:66, SEQ ID NO:68 and SEQ ID NO:77. This has the effect of altering monoterpene biosynthesis, thereby increasing the production of monoterpenes, as well as providing novel seed oils having desirable monoterpene compositions. In this manner, the transformed seed provides a factory for the production of modified oils. The modified oil itself may be used and/or the compounds in the oils can be isolated. Thus, the present invention allows for the production of particular monoterpenes of interest as well as speciality oils.

The nucleic acid encoding the monoterpene synthases of the present invention can be used in expression cassettes for expression in the transformed plant tissues. To alter the monoterpene levels in a plant of interest, the plant is transformed with at least one expression cassette comprising a transcriptional initiation region linked to a nucleic acid sequence encoding a monoterpene synthase. Such an expression cassette is provided with a plurality of restriction sites for insertion of the nucleic acid sequence encoding a monoterpene synthase so that it is under the transcriptional regulation of the regulatory regions.

The transcriptional initiation sequence may be native or analogous to the host or foreign or heterologous to the host. In this regard, the term "foreign" means that the transcriptional initiation sequence is not found in the wild-type host into which the transcriptional initiation region is introduced.

Of particular interest are those transcriptional initiation regions associated with storage proteins, such as napin, cruciferin, β-conglycinin, phaseolin, globulin or the like, and proteins involved in fatty acid biosynthesis, such as acyl carrier protein (ACP). See, U.S. Patent No. 5,420,034, herein incorporated by reference.

The transcriptional cassette will preferably include, in the 5' to 3' direction of transcription, a transcriptional and translational initiation region, a monoterpene synthase DNA sequence of interest, and a transcriptional and translational termination region functional in plants. The termination region may be from the same organism as the transcriptional initiation region, may be from the same organism as the

monoterpene synthase DNA, or may be derived from another source. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. Other termination sequences are set forth in Guerineau et al., (1991), *Mol. Gen. Genet.*, **262**:141-144; Proudfoot, (1991), *Cell*, **64**:671-674; Sanfacon et al., (1991), *Genes Dev.*, **5**:141-149; Mogen et al., (1990), *Plant Cell*, **2**:1261-1272; Munroe et al., (1990), *Gene*, **91**:151-15 8; Ballas et al., (1989), *Nucleic Acids Res.*, **17**:7891-7903; Joshi et al., (1987), *Nucleic Acid Res.*, **15**:9627-9639).

In the presently preferred form of the invention, a nucleic acid sequence encoding a monoterpene synthase protein will be targeted to plastids, such as chloroplasts, for expression. Thus, the nucleic acid sequence, or sequences, encoding a monoterpene synthase protein, or proteins, may be inserted into the plastid for expression with appropriate plastid constructs and regulatory elements. Alternatively, nuclear transformation may be used in which case the expression cassette will contain a nucleic acid sequence encoding a transit peptide to direct the monoterpene biosynthesis enzyme of interest to the plastid. Such transit peptides are known in the art. See, for example, Von Heijne et al. (1991) *Plant Mol. Biol. Rep.* **9**:104-126; Clark et al. (1989) *J. Biol. Chem.* **264**:17544-17550; della-Cioppa et al. (1987) *Plant Physiol.* **84**:965-968; Romer et al. (1993) *Biochem. Biophys. Res. Commun.* **196**:1414-1421; and, Shah et al. (1986) *Science* **233**:478-481. Nucleic acid sequences encoding monoterpene synthases of the present invention may utilize native or heterologous transit peptides.

The construct may also include any other necessary regulators such as plant translational consensus sequences (Joshi, C.P., (1987), *Nucleic Acids Research*, **15**:6643-6653), introns (Luehrsen and Walbot, (1991), *Mol. Gen. Genet.*, **225**:81-93) and the like, operably linked to a nucleotide sequence encoding a monoterpene synthase of the present invention.

It may be beneficial to include 5' leader sequences in the expression cassette which can act to enhance translation. Translation leaders are known in the art and include: picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5' noncoding region) (Elroy-Stein, O., Fuerst, T.R., and Moss, B. (1989) *PNAS USA* **86**:6126-6130); potyvirus leaders, for example, TEV leader (Tobacco Etch Virus) (Allison et al. (1986); MDMV leader (Maize Dwarf Mosaic Virus); *Virology*, **154**:9-20), and human immunoglobulin heavy-chain binding protein (BiP), (Macejak, D.G., and Sarnow, P. (1991), *Nature*, **353**:90-94; untranslated leader from the coat protein

mRNA of alfalfa mosaic virus (AMV RNA 4), (Jobling, S.A., and Gehrke, L. (1987), *Nature*, **325**:622-625; tobacco mosaic virus leader (TMV), (Gallie, D.R. et al. (1989), *Molecular Biology of RNA*, pages 237-256; and maize chlorotic mottle virus leader (MCMV) (Lommel, S.A. et al. (1991), *Virology*, **81**:382-385. See also, Della-Cioppa et al., (1987), *Plant Physiology*, **84**:965-968.

Depending upon where the monoterpene synthase sequence of interest is to be expressed, it may be desirable to synthesize the sequence with plant preferred codons, or alternatively with chloroplast preferred codons. The plant preferred codons may be determined from the codons of highest frequency in the proteins expressed in the largest amount in the particular plant species of interest. See, EPA 0359472; EPA 0385962; WO 91/16432; Perlak et al. (1991) *Proc. Natl. Acad. Sci. USA* **88**:3324-3328; and Murray et al. (1989) *Nucleic Acids Research* **17**:477-498. In this manner, the nucleotide sequences can be optimized for expression in any plant. It is recognized that all or any part of the nucleic acid sequence encoding a monoterpene synthase protein may be optimized or synthetic. That is, synthetic or partially optimized sequences may also be used. For the construction of chloroplast preferred genes, see U.S. Patent No. 5,545,817.

In preparing the transcription cassette, the various DNA fragments may be manipulated, so as to provide for the DNA sequences in the proper orientation and in the proper reading frame. Towards this end, adapters or linkers may be employed to join the DNA fragments or other manipulations may be involved to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like. For this purpose, *in vitro* mutagenesis, primer repair, restriction, annealing, resection, ligation, or the like may be employed, where insertions, deletions or substitutions, such as transitions and transversions, may be involved.

The recombinant DNA molecules of the invention can be introduced into the plant cell in a number of art-recognized ways. Those skilled in the art will appreciate that the choice of method might depend on the type of plant, *i.e.*, monocot or dicot, targeted for transformation. Suitable methods of transforming plant cells include microinjection (Crossway et al. (1986) *BioTechniques* **4**:320-334), electroporation (Riggs et al. (1986) *Proc. Natl. Acad. Sci. USA* **83**:5602-5606), *Agrobacterium* mediated transformation (Hinchee et al. (1988) *Biotechnology* **6**:915-921) and ballistic particle acceleration (see, for example, Sanford et al., U.S. Patent No. 4,945,050; and McCabe et al. (1988) *Biotechnology* **6**:923-926). Also see, Weissinger et al. (1988) *Annual Rev. Genet.* **22**:421-477; Sanford et al. (1987)

Particulate Science and Technology **5**:27-37 (onion); Christou et al. (1988) *Plant Physiol.* **87**:671-674 (soybean); McCabe et al. (1988) *Bio/Technology* **6**:923-926 (soybean); Datta et al. (1990) *Biotechnology* **8**:736-740 (rice); Klein et al. (1988) *Proc. Natl. Acad Sci. USA*, **85**:4305-4309 (maize); Klein et al. (1988) *Biotechnology* **6**:559-563 (maize); Klein et al. (1988) *Plant Physiol.* **91**:440-444 (maize); Fronun et al. (1990) *Biotechnology* **8**:833-839; and Gordon-Kamm et al. (1990) *Plant Cell* **2**:603-618 (maize).

Alternatively, a plant plastid can be transformed directly. Stable transformation of chloroplasts has been reported in higher plants, see, for example, SVAB et al. (1990) *Proc. Nat'l. Acad. Sci. USA* **87**:85268530; SVAB & Maliga (1993) *Proc. Natl. Acad. Sci. USA* **90**:913-917; Staub & Maliga (1993) *Embo J.* **12**:601-606. The method relies on particle gun delivery of DNA containing a selectable marker and targeting of the DNA to the plastid genome through homologous recombination. In such methods, plastid gene expression can be accomplished by use of a plastid gene promoter or by trans-activation of a silent plastid-borne transgene positioned for expression from a selective promoter sequence such as that recognized by T7 RNA polymerase. The silent plastid gene is activated by expression of the specific RNA polymerase from a nuclear expression construct and targeting of the polymerase to the plastid by use of a transit peptide. Tissue-specific expression may be obtained in such a method by use of a nuclear-encoded and plastid-directed specific RNA polymerase expressed from a suitable plant tissue specific promoter. Such a system has been reported in McBride et al. (1994) *Proc. Natl. Acad Sci. USA* **91**:7301-7305.

The cells which have been transformed may be grown into plants by a variety of art-recognized means. See, for example, McConnick et al., *Plant Cell Reports* (1986), **5**:81-84. These plants may then be grown, and either selfed or crossed with a different plant strain, and the resulting homozygotes or hybrids having the desired phenotypic characteristic identified. Two or more generations may be grown to ensure that the subject phenotypic characteristic is stably maintained and inherited and then seeds harvested to ensure the desired phenotype or other property has been achieved.

As a host cell, any plant variety may be employed. Of particular interest, are plant species which provide seeds of commercial value. For the most part, plants will be chosen where the seed is produced in high amounts, a seed-specific product of interest is involved, or the seed or a seed part is edible. Seeds of interest in the

practice of the present invention include, but are not limited to, the oil seeds, such as oilseed Brassica seeds, cotton seeds, soybean, safflower, sunflower, coconut, palm, and the like; grain seeds such as wheat, barley, oats, amaranth, flax, rye, triticale, rice and corn; other edible seeds or seeds with edible parts including pumpkin, squash, sesame, poppy, grape, mung beans, peanut, peas, beans, radish, alfalfa, cocoa, and coffee; and tree nuts such as walnuts, almonds, pecans, and chick-peas.

While the preferred embodiment of the invention has been illustrated and described, it will be appreciated that various changes can be made therein without departing from the spirit and scope of the invention.